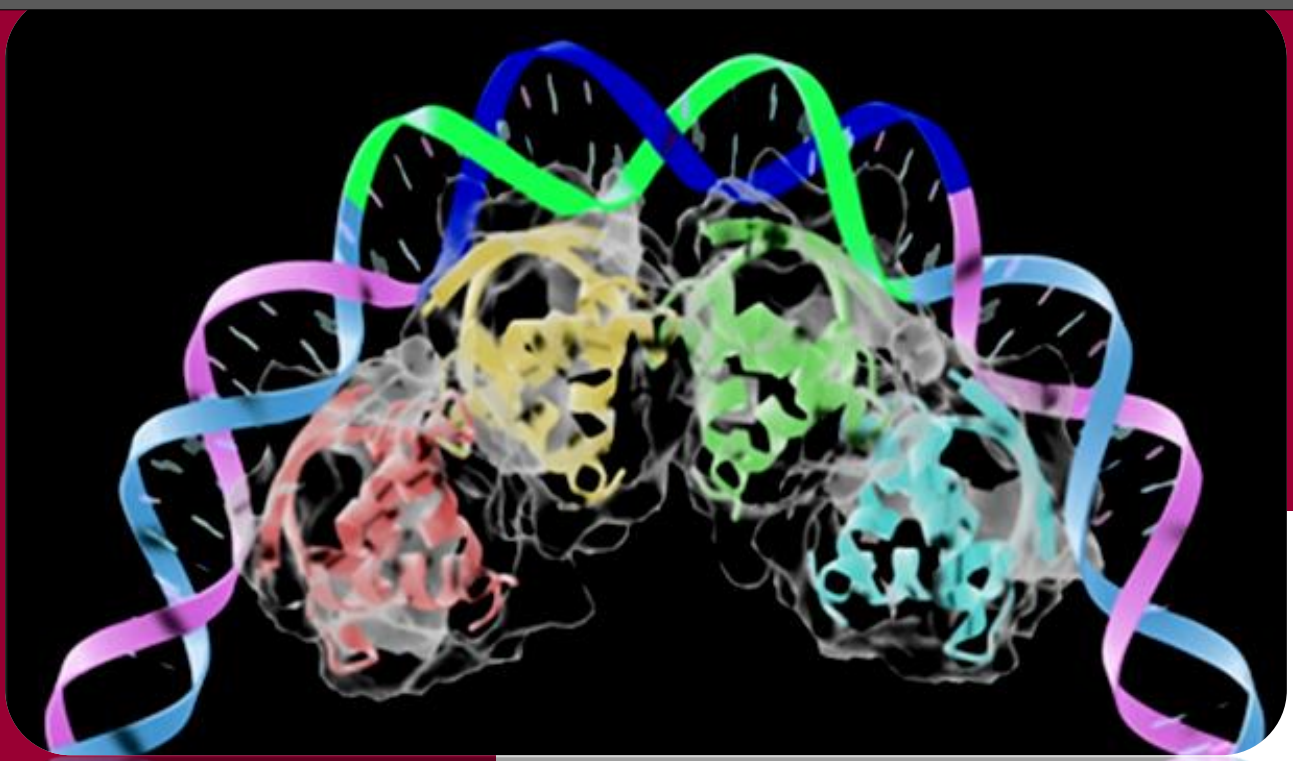
UNIVERSIDAD AUTONOMA DE MADRID

Faculty of Sciences

Department of Molecular Biology

Analysis of the cooperative interactions between CopG dimers bound to subsites of its operator DNA



PhD Thesis
by

TANIA SAMIR RUBIO LEPE

Madrid, 2013

Faculty of Sciences

Department of Molecular Biology

Analysis of the cooperative interactions between CopG dimers bound to subsites of its operator DNA

by

Tania Samir Rubio Lepe

Dissertation submitted for the Degree of **Doctor of Philosophy** by the Universidad Autónoma de Madrid in the Molecular Biology PhD Program

Supervised by

Gloria del Solar Dongil, PhD
Centro de Investigaciones Biológicas
CSIC

Madrid, 2013

The financial support for this research was granted by the Spanish Ministerio de Ciencia e Innovación (MICINN). The work was performed in the Centro de Investigaciones Biológicas (CIB) (Grants: BFU2007-63575, BFU2010-19597) and in the Center for Cooperative Research in Biosciences (CIC bioGUNE) (Grant: CTQ2009-10353/BQU). I was a recipient of a PreDoc JAE fellowship (JAEpre_07_00437) by the Consejo Superior de Investigaciones Científicas (CSIC).

Front: 'Model of CopG binding to its target DNA'. Taken and modified from (del Solar, G. *et al*. 2002).

For my family and José

Thanks to everybody who held in their minds that this thesis would be a reality.

And specially thanks to José for not letting me give up…for his constant encouragement and financial support, otherwise, this book wouldn't have been written.

Contents

Contents

Page

FIGURES

Introduction **Page**

Materials and Methods

Results

CHAPTER I: Binding of repressor CopG to its operator DNA

CHAPTER II: Structural features of three CopG variants

CHAPTER III: Interactions of the three CopG variants with DNA

CHAPTER IV: Super-operators

Discussion

TABLES

Materials and Methods **Page**

Chapter I

Chapter II

Chapter III

Chapter IV

Abbreviations

Å	angstrom
aa	amino acid (s)
Amp	ampicillin
APS	ammonium persulfate
ATP	adenosine triphosphate
bp	base pair (s)
BSA	bovine serum albumin
°C	degrees Celsius
CD	circular dichroism
Ci	curies
Cm	chloramphenicol
cpm	counts per minute
CV	column volume
DMS	dimethyl-sulfate
DNAse I	desoxirribonuclease I
dsDNA	double-stranded DNA
dso	double-strand origin
DTT	dithiothreitol
EDTA	ethylenediamine-tetra-acetic acid
EMSA	electrophoretic mobility shift assay
Eq	equation
EtBr	ethidium bromide
ext.	extended
FPLC	fast protein liquid chromatography
h	hour (s)
HPLC	high-performance liquid chromatography
IMAC	immobilized metal affinity chromatography
IPTG	isopropyl-β-D-thiogalactosidase
°K	degrees Kelvin
kb	kilobase (s)
kDa	kilo dalton (s)
kJ	kilo joule (s)
Km	kanamycin
kpsi	kilo pound square per inch
M&M	Materials and Methods
min	minute (s)
N	plasmid copy number
nm	nanometer (s)
NMR	nuclear magnetic resonance
NOE	nuclear Overhausser effect
nt	nucleotide (s)
OD	optical density
OH•	hydroxyl radical
PAA	polyacrylamide

PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PDB	Protein Data Bank
ppm	parts per milion
pRNA	RNA primer
R.T.	room temperature
RMSD	root mean square deviation
RNAP	RNA polymerase
rpm	revolutions per minute
s	seconds
SD	standard deviation
SDS	sodium dodecyl sulfate
SE	Symmetric Element
ssDNA	single-stranded DNA
sso	single-strand origin
T4-PNK	polynucleotide kinase of T4-Bacteriophage
TAE	tris-acetate-EDTA
TBE	tris-borate-EDTA
Tc	tetracyclin
TEMED	N,N,N',N'-tetra-methylethylenediamine
Tris	tris-hydroxymethyl-aminomethane
UV	ultraviolet
V	volts
v/v	volume/volume
W	watts

Abstract/Resumen

Analysis of the cooperative interactions between CopG dimers bound to subsites of its operator DNA

ABSTRACT

Protein CopG (45 aa) encoded by plasmid pMV158 is the smallest transcriptional repressor characterized so far, and is the prototype of a family of plasmids replicating by the Rolling-Circle mechanism. CopG represses P_{cr} promoter, which directs the transcription of its own gen (*copG*) and that of the initiator protein for plasmid replication (*repB*). The DNA target of CopG has a symmetric element (SE) that includes a repeat of the sequence 5'-TGCA-3' at each side of the symmetry axis, in the subsites called the Left and Right Symmetric Elements (LSE and RSE, respectively). The LSE overlaps the -35 promoter region. Crystallographic studies of CopG free and bound to the SE demonstrated that it has a "β-Ribbon-α1-Helix-α2-Helix" (RHH) three-dimensional configuration, and that one dimer of protein binds to each half of the SE. The CopG-DNA interaction relies on the contacts made by the β-Ribbon with DNA bases, and on those made by the α2 helix N-terminal residues with the phosphate backbone. Binding of CopG to its target DNA generates, with a pattern of cooperativity, four well-defined complexes (CI, CII, CIII and CIV), which are supposed to correspond to 1, 2, 3 and 4 dimers bound to the operator DNA. Thus, one dimer binds to each subsite of the SE, and two extra dimers would bind laterally to the SE, in regions located one helical turn apart, defined as the Left and Right Arms (LA and RA, respectively). The RA overlaps the extended -10 region of P_{cr} promoter.

In this Thesis we demonstrated that, in fact, the specific nucleoprotein complex formed by CopG consists of four dimers bound in a sequential and highly cooperative manner to four subsites of the DNA spanning about 50 bp. We also determined that the RSE is the primary site of protein binding, to which CopG binds establishing interactions beyond those described in the SE-CopG tetramer co-crystals. Then, the second dimer binds to the RA, the third to the LSE and the last one binds to the LA. These results support the hypothesis about the CopG-mediated transcriptional repression by active dissociation, according to which the RSE keeps accessible for CopG binding once the RNAP is bound to the P_{cr} promoter. Binding of CopG to the RSE then promotes, by cooperative interactions, the formation of the complete nucleoprotein complex, dislodging finally the RNAP. On the other hand, two CopG variants were studied: $CopG_{A30E}$ and $CopG_{G25E}$. The altered residues in these mutants are located in the dimer-dimer interaction surface observed in the SE-CopG tetramer co-crystals. By NMR, differences in the structural configuration of the three CopG variants were not observed. However, EMSAs revealed a diminished affinity for DNA and altered cooperative properties in both mutants compared to the wild-type protein. Using the three CopG variants and a battery of operator mutants, we observed and proposed that: i) the recognition sequence that better promotes binding of CopG was the 6-bp motif 5'-(R)TGCA(Y)-3', and ii) upon binding, the four dimers arrange as two tetramers, each bound to a half of the operator, hence generating at least two different dimer-dimer interfaces in the whole nucleoprotein complex formed by CopG, namely the interaction surface observed for the two dimers bound to the SE as described in the co-crystals, and the interface generated between the two dimers bound to either of the operator halves. All together, we demonstrated that binding of CopG to its operator DNA depends both on the nucleotide sequence and on the cooperative interdimeric interactions.

Análisis de las interacciones cooperativas entre dímeros de CopG unidos a los subsitios de su operador

RESUMEN

La proteína CopG (45 aa) codificada por el plásmido pMV158 es el represor transcripcional más pequeño descrito hasta ahora y es el prototipo de una familia de plásmidos que replican por el mecanismo de círculo rodante. CopG reprime el promotor P_{cr}, el cual dirige la transcripción tanto de su propio gen (*copG*) como del de la proteína iniciadora de la replicación (*repB*). La diana de CopG posee un elemento simétrico (SE) de 13 pb, que presenta una repetición de la secuencia 5'-TGCA-3' en los subsitios denominados Elemento Simétrico Izquierdo y Derecho (LSE y RSE, respectivamente), a cada lado del eje de simetría. El LSE solapa con la caja -35 del promotor P_{cr}. Estudios cristalográficos de CopG libre y unida al SE demostraron que posee una estructura del tipo "β-Ribbon-α1-Helix-α2-Helix" (RHH), y que un dímero de proteína se une a cada mitad del SE. La interacción CopG-DNA reside en los contactos establecidos por la lámina β antiparalela (Ribbon) con las bases de DNA, y los contactos de los residuos del extremo N-terminal de la α-hélice 2 con los grupos fosfato. La unión de CopG a su operador genera, con un patrón de alta cooperatividad, cuatro complejos bien definidos (CI, CII, CIII y CIV), que pensamos se corresponderían, respectivamente, a 1, 2, 3 y 4 dímeros de proteína unidos al operador. Así, dos dímeros se unen a los subsitios LSE y RSE, y dos dímeros más se unirían a una vuelta de hélice de cada subsitio del SE, en las regiones llamadas brazo izquierdo (LA) y brazo derecho (RA). El RA solaparía con la caja -10 extendido del promotor P_{cr}.

En esta Tesis demostramos que, efectivamente, el complejo nucleoproteico específico formado por CopG consiste en 4 dímeros de proteína unidos, de forma secuencial y altamente cooperativa, a 4 sitios del DNA que abarcan aproximadamente 50 pb. Demostramos que el sitio primario de unión es el RSE, al cual se une el primer dímero, formando el CI. El segundo dímero se une al RA, el tercero al LSE y el cuarto, al LA. Estos resultados sustentan la hipótesis sobre el mecanismo de disociación activa del complejo abierto mediado por CopG, según la cual, una vez unida la RNAP al promotor P_{cr}, el RSE permanece accesible para la unión de CopG, lo que promueve, mediante interacciones cooperativas, la formación del complejo nucleoproteico entero, consiguiendo así desplazar a la enzima. Por otro lado, se estudiaron dos mutantes del represor: $CopG_{A30E}$ y $CopG_{G25E}$. Los residuos alterados en estas variantes proteicas se localizan en la superficie de interacción dímero-dímero observada en los co-cristales de CopG unida al SE. No se observaron diferencias en la configuración estructural de las tres variantes proteicas; sin embargo, ensayos de unión a DNA revelaron una disminución en la afinidad y una alteración significativa de las propiedades cooperativas de ambos mutantes con respecto a la proteína silvestre. Usando las tres proteínas y diversos operadores mutantes para la unión de CopG, observamos y proponemos que: i) la secuencia nucleotídica óptima para la unión de CopG sería 5'-(R)TGCA(Y)-3', y ii) durante su unión, los cuatro dímeros se disponen como dos tetrámeros cada uno unido a una mitad del operador, generando por tanto, al menos dos superficies de interacción distintas en el complejo entero: una, la generada entre los 2 dímeros unidos al SE, y otra, la formada entre los 2 dímeros (que integran el tetrámero) de CopG, unidos a cada una de las mitades del operador entero. En conjunto, demostramos que la unión de CopG a su DNA operador depende tanto de la secuencia nucleotídica como de las interacciones cooperativas dímero-dímero.

Introduction

1. Plasmids

Plasmids are extrachromosomal double-stranded (ds) DNA elements that have been found in species from the three major domains of the living world, *Archaea*, *Bacteria* and *Eucarya*, and are involved in the spread of genetic information within the environment. The characteristic that better defines plasmids is that they replicate in an autonomous and self-controlled way, in coordination with cellular division. Natural plasmids vary considerably in size (from a few hundreds of bp to hundreds of kb), in copy number (from 1 to about 30), in host range (from one to several species), and in the various genetic traits conferred to their hosts [reviewed in (Espinosa, Cohen *et al.* 2005)]. Plasmids constitute from 1% to greater than 15% of the genome of many bacterial species, and maintain themselves stable among bacterial populations throughout vertical and horizontal transference mechanisms.

The phenotypic characters specified by plasmids include antibiotic resistance, toxic heavy metal resistance, degradation of xenobiotic compounds, symbiotic and virulence determinants, bacteriocin production, resistance to radiation, increased mutation frequency and even transfer of DNA to higher eukaryotes. Although genetic information encoding these traits could be considered dispensable, they mostly offer evolutionary advantages to the host.

The wealth of genetic information carried by plasmids, their impact in the microbial communities, and the potential of these elements to act as natural cloning vectors have stimulated research into plasmids not only from the fundamental but also from the clinical, biotechnological, and environmental points of view.

1.1. Plasmid replication mechanisms

The genes and sites required for autonomous replication and its control constitute the basic replicon of the plasmids. They generally consist of: a) the origin(s) of replication (generically termed *ori*), which is characteristic of each replicon; b) the genes encoding Rep proteins, involved in the initiation of replication and, c) the genes responsible for the control of replication. After assembling of the replisome, plasmids employ different mechanisms to carry out their replication, namely, theta type, strand displacement and rolling circle [reviewed in (del Solar, Giraldo *et al.* 1998)].

1.1.1. Theta type

This mechanism is widespread among plasmids from Gram-negative bacteria, although it has also been described for plasmids isolated from Gram-positive ones. Its name derives from the

fact that under electron microscopy (EM), the replication intermediates are seen as typical Θ ("theta")-shaped molecules. In brief, DNA replication by theta mechanism involves melting of the parental strands, synthesis of a primer RNA (pRNA), and initiation of DNA synthesis by covalent extension of the pRNA. Opening of the strands is catalyzed by specific initiators (Rep and DnaA proteins) and/or by transcription by RNA polymerase (RNAP). Initiation proteins promote, at the origin of replication, the sequential assembly of components of the replisome complex. The main replicative helicase of the cell catalyzes further unwinding of the strands. DNA synthesis of both strands is coupled and occurs continuously on one of them (leading-strand) and discontinuously on the other (lagging-strand). DNA polymerase III is required for elongation of plasmid DNA replication. In addition, DNA polymerase I can participate in the early synthesis of the leading strand (ColE1 and pAMβ1). Theta-type replication is, in most cases, unidirectional. Topoisomers are originated at termination (right-handed catenates), and their resolution requires the participation of Topoisomerase IV. Termination of DNA replication is determined in some plasmids by specific protein-DNA complexes.

1.1.2. Strand displacement

The best-known examples of plasmids replicating by this mechanism are the promiscuous plasmids of the IncQ family, whose prototype is RSF1010. Replication of this plasmid occurs from two symmetrical and adjacent single-stranded initiation sites (*ssiA* and *ssiB*) positioned one on each DNA strand. Replication starts when these origins are exposed as single-stranded regions. The melting of the DNA strand is dependent on two plasmid replication proteins, RepC and RepA, and is facilitated by an AT-rich region that precedes the *ssiA* and *ssiB* regions. RepC recognizes directly repeated sequences of the origin adjacent to the AT-rich region, and RepA is a DNA helicase. Priming of DNA synthesis at these origins is catalyzed by the plasmid-specific primase (RepB). Synthesis of each one of the strands occurs continuously and results in the displacement of the complementary strand. Replication of this displaced strand is initiated at the exposed single-stranded initiation site. Due to the activities of the three plasmid replication proteins (RepA, RepB and RepC), initiation of RSF1010 replication is independent of transcription by host RNAP and of host replication factors acting at the early replication stages (DnaA, DnaB, DnaC, and DnaG). This independence may account for the broad-host range character of the IncQ replicons.

1.1.3. Rolling-Circle replication

As concerning with this thesis, this mechanism will be detailed.

Replication by the Rolling-Circle (RC) mechanism was discovered in single-stranded DNA (ssDNA) bacteriophages of *Escherichia coli*, and afterwards analyzed in bacterial plasmids (Khan, Adler *et al.* 1982; Koepsel, Murray *et al.* 1985; del Solar, Moscoso *et al.* 1993). Some plant-mitochondrial plasmids and Parvoviridae (animal-viruses) and Geminiviridae (plant-viruses) families also replicate by this process (Berns 1990; Backert, Meissner *et al.* 1997).

Bacterial RC-replicating (RCR) plasmids were discovered many years ago (Khan, Adler *et al.* 1982; Koepsel, Murray *et al.* 1985; te Riele, Michel *et al.* 1986) and initially were thought to be an exception to the rule that most plasmids replicate by a theta-type mechanism. Nowadays, more than 200 RCR-plasmids have been identified making it possible to analyze their phylogenetic relationships and demonstrate that they come from a few common ancestors. RCR-plasmids are mostly abundant in Gram-positive bacteria, but they also have also been isolated from many Gram-negative hosts, cyanobacteria and species of *Archaea* (Gielow, Diederich *et al.* 1991; Yasukawa, Hase *et al.* 1991; Erauso, Marsin *et al.* 1996; Rozhon, Khan *et al.* 2011).

The RC replication is unidirectional and considered to be asymmetric because synthesis of the leading and the lagging DNA strands is uncoupled. One of the most relevant features of RC replication is that the newly synthesized leading strand remains covalently bound to the same parental strand. All RCR-plasmids encode initiator proteins that have origin-specific binding and nicking-closing activities. The double-strand origin (*dso*) contains both the binding (*bind*) and the nicking (*nic*) loci for the Rep proteins. Initiation of replication requires both loci, while termination of plasmid RC replication can be promoted by a smaller region containing only the nick sequence. The single-strand origin (*sso*) of RCR-plasmids is important for the conversion of the displaced parental leading-strand to the double-stranded form. The current model for plasmid RC-replication is based on studies with plasmids of the pT181 family (Figure 1), and can be summarized as follows [reviewed in (Khan 2005)].

The Rep protein interacts with its specific *bind* sequence located within the plasmid *dso*. The Rep-*dso* interaction may result in a sharp bend in the DNA and/or generation of a hairpin in which the Rep nick sequence is located in a single-stranded loop. The Rep protein then nicks the *dso* at the nick site and becomes covalently attached to the 5'-phosphate through a tyrosine residue present in its active site. Rep proteins also recruit a bacterial DNA helicase, through a specific protein-protein interaction. The helicase then unwinds the DNA, and the ssDNA binding (SSB) protein coats the displaced ssDNA. Leading-strand replication initiates by extension synthesis at the free 3'-OH end at the nick site by DNA polymerase III and proceeds until the parental leading-strand has been fully displaced. Once the replisome reaches the termination

site, i.e. the regenerated *dso*, DNA synthesis continues to approximately 10 nucleotides (nt) beyond the Rep nick site. The second, free monomer of the dimeric Rep protein then cleaves the displaced ssDNA. Following a series of cleavage/rejoining events, the parental leading-strand is released as a circular ssDNA replicative-intermediate, along with a relaxed, closed circular dsDNA containing the newly synthesized leading-strand. The dsDNA is then supercoiled by the host DNA gyrase. The Rep protein, in which the active tyrosine site of one protomer is covalently attached to the 10-mer oligonucleotide, is released. This form of Rep, termed RepC/RepC* for pT181, which has catalyzed one round of leading strand replication is inactive in further replication. The ssDNA released after completion of the leading-strand synthesis is converted to dsDNA utilizing the *sso* and host proteins. This involves the synthesis of a pRNA by the RNAP, which primes replication around the circle by DNA polymerase III. Finally, the DNA ends are joined by DNA ligase and the resultant dsDNA is converted to the supercoiled (SC) form by DNA gyrase.

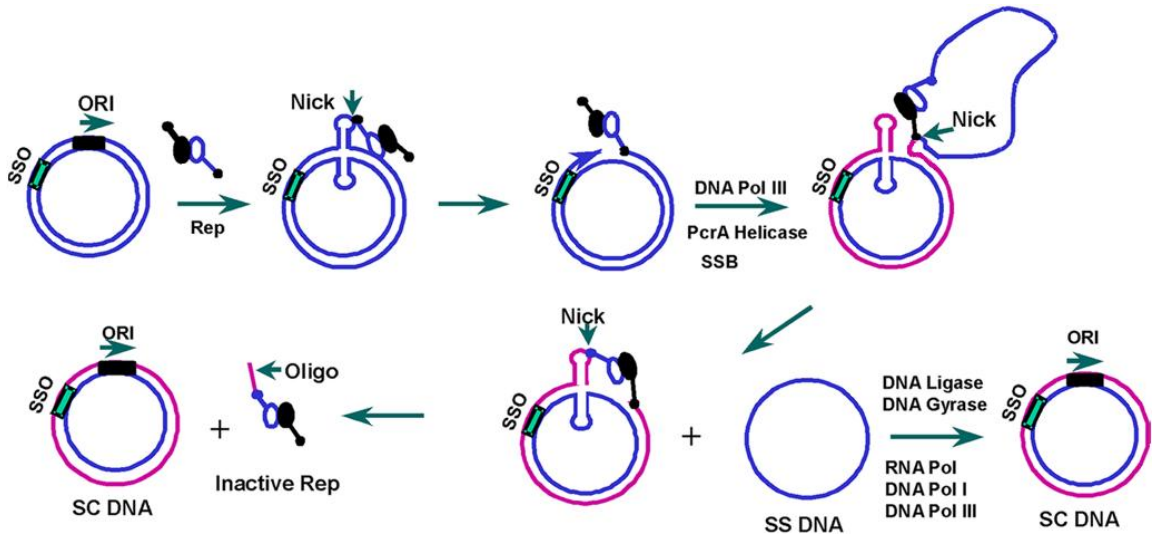


Figure 1. A model for plasmid RC replication. Rep protein recognizes the origin of replication (*dso*) on the supercoiled DNA and generates a 3'-OH end. From here, synthesis of the leading strand is initiated. When the replisome reaches a *dso* reconstituted, Rep catalyzes a reaction of trans-esterification that releases the molecule of dsDNA and a replicative-intermediate of ssDNA. From the origin of replication (*sso*) of the lagging strand, which is exposed in the ssDNA molecule, a second copy of plasmid dsDNA is generated. Finally, DNA ends are joined by the ligase, and the gyrase converts the dsDNA molecule in its supercoiled (SC) form. Taken and modified from (Khan 2005). Details are given in the text.

RCR-plasmids are usually less than 10 kb in size, have a high copy number per cell and are tightly organized; suggesting that economy of genetic information is an important aspect of these plasmids. All RCR-replicons described so far can be divided into more than a dozen families based on sequence homologies in the initiator Rep protein and *dso* (Khan 2005). However, based on the genetic elements of replication and control (del Solar, Moscoso *et al.* 1993), four main

plasmid families have been defined, and are represented by pT181, pUB110/pC194, pMV158, and pSN2 plasmids (Gruss & Ehrlich 1989; Novick 1989; del Solar, Moscoso *et al.* 1993).

1.2. Control of plasmid replication

A particular plasmid in a given host under steady growth conditions is kept with a characteristic average number of copies, which allows it to co-exist stably with, and minimize the metabolic burden to the host. For that, plasmids must control their replication and maintain its defined copy number (*N*) inside the cells.

Two different stages of the plasmid life cycle have been defined. Firstly, the "establishment" that occurs when a plasmid copy enters into a new permissive host; the successful establishment may depend upon the plasmid ability to replicate quickly before the host divides (Highlander & Novick 1987). This may result in an overshoot in *N* before it reaches the characteristic average value (N_{av}). In the second stage, the replicon enters into the steady state in which the copy number is constant, maintaining in average one replicative event per plasmid copy and cell cycle (Nordström & Wagner 1994). To keep this average frequency, plasmids use self-encoded negative control systems that are able to "sense" and correct up and down fluctuations from N_{av} in individual cells. These control systems adjust the rate of replication per plasmid copy, so that it becomes higher or lower than 1, depending on whether there has been a decrease or an increase, respectively, in copy number with respect to the N_{av} in a given cell.

Control of replication by negative regulators and random selection of plasmids for replication has an additional consequence: two different plasmids with identical replicons are unable to stably co-exist within a given cell in the absence of selective pressure. This leads to segregation of plasmids within the host population, a phenomenon known as plasmid incompatibility. Control of replication by inhibitors would require that they could measure the concentration of plasmid copies within the cell. This could be achieved by an unstable inhibitor synthesized constitutively or by a stable inhibitor synthesized shortly after each initiation event (Pritchard, Barth *et al.* 1969). Any of these alternatives would increase or decrease the rate of initiation of replication when the average copy number is, respectively, lower or higher than required. When the frequency of initiation is determined by the level of an initiator protein, controlled plasmid replication would include the inactivation of the initiator protein after each replication event (Rasooly & Novick 1993).

Out of the three stages of plasmid replication (initiation, elongation and termination), all plasmids control their copy number at the initiation step by specifying a negative feedback loop

(Pritchard, Barth *et al.* 1969). This regulation is exerted by some strategies: i) small antisense-RNAs that hybridize complementarily with essential RNAs (pRNA or *rep* mRNA); ii) direct-repeated sequences (iterons) that can interact with the Rep protein, and have influence on its auto-regulation, or even they can deactivate the origin of replication by the *handcuffing* mechanism (Chattoraj 2000); and iii) the joint action of a transcriptional repressor and a counter-transcribed RNA (ctRNA) (del Solar & Espinosa 2000). Within the last group, there are two categories. In one of them, the ctRNA plays the main regulatory role, whereas the protein has been proposed as only an auxiliary element. And in the second one, both elements acting on different targets efficiently correct fluctuations in the *N*-value at the steady state (del Solar, Acebo *et al.* 1995; del Solar, Giraldo *et al.* 1998).

2. Plasmid pMV158

Promiscuous plasmid pMV158 (5536 bp) was initially isolated from *Streptococcus agalactiae* (Burdett 1980) and it has been studied for more than 20 years in our lab. Nowadays, it is considered the prototype of a family of RCR-plasmids comprising more than 60 plasmids isolated from different bacteria. Plasmid pMV158 is able to colonize distinct species, among them, *S. pneumoniae*, *Staphylococcus aureus*, *Lactococcus lactis, Enterococcus faecalis*, *Bacillus subtilis* and *E. coli.* It has been demonstrated that pMV158 is mobilized by auxiliary plasmids (like pIP501) among strains of *S. pneumoniae*, or transferred to other species as *E. faecalis* and *L. lactis* (Smith, Shoemaker *et al.* 1980; Farías & Espinosa 2000; Nieto & Espinosa 2003; Lorenzo-Díaz & Espinosa 2009).

Genetic structure of pMV158 is organized into functional cassettes: i) the LIC (Leading strand Initiation and Control) module, which contains the *dso* and the genes essential for replication and its control; ii) the DET module, that provides the antibiotic (tetracycline)-resistance determinant; iii) the MOB module, which is involved in plasmid mobilization and interplasmidic recombination (Novick 1989; Priebe & Lacks 1989); and iv) two *sso* that flank the MOB module (Figure 2). Some pMV158 family members lack one or several cassettes, but the LIC module is essential (del Solar, Moscoso *et al.* 1993).

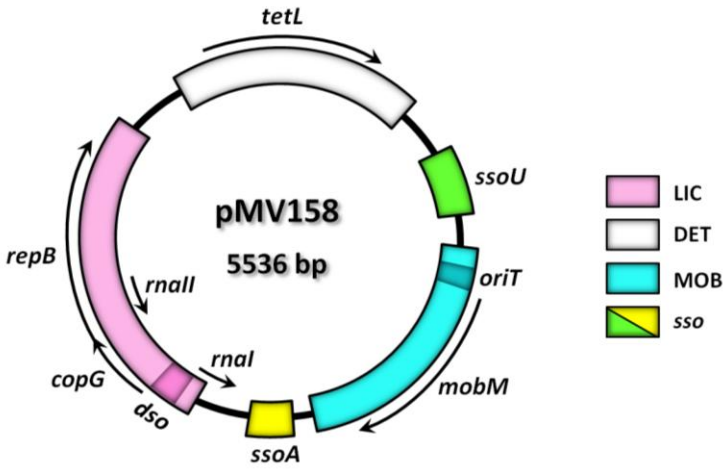


Figure 2. Functional and genetic map of pMV158. Genes harbored by the plasmid are indicated. The *copG-repB* operon gives rise to the replication initiator protein (RepB) and to the transcriptional repressor (CopG). The product of *tetL* confers tetracycline-resistance. Gene *mobM* codifies the protein responsible for initiation of the plasmid conjugative mobilization. The product from *rnaII* is involved in control of replication at a translational level, and *rnaI* has not a defined function. Direction of transcription is indicated by *arrows*. The scheme also shows the *dso* and *sso* origins, from which replication of the leading and lagging strands, respectively, is initiated.

2.1. Control of pMV158 replication

In general, control of replication in RCR-plasmids is carried out by a tight control on the synthesis of the replication initiator RepB protein; in the case of pMV158, such a control involves the joint action of a ctRNA and a transcriptional repressor Cop protein (del Solar & Espinosa 1992; del Solar, Acebo *et al.* 1995). This dual regulation has also been shown for the theta-replicating plasmids of the pIP501 family (Brantl 1994).

The characterization of some pMV158 mutations that resulted in increased *N*-values along with the definition of DNA regions that affected plasmid replication *in trans* (*inc* determinants), allowed the establishment of a novel mechanism of replication control involving transcriptional repressor CopG and antisense RNAII; both implicated in regulation of synthesis of the initiator RepB protein (del Solar, Acebo *et al.* 1995).

Protein CopG binds to and represses the transcription from the single promoter (P_{cr}) that directs the synthesis of a bicistronic mRNA for both CopG and RepB (Figure 3). Thus, by binding to the *copG-repB* promoter region, CopG regulates its own synthesis and that of RepB (del Solar, Pérez-Martín *et al.* 1990). Mutations or deletions in *copG* lead to an increase in the plasmid copy number (Acebo, Alda *et al.* 1996), while *copG* exerts only a weak incompatibility toward plasmids with the pMV158 replicon when cloned, under its own transcription and translation signals, into a high-copy number compatible vector (del Solar & Espinosa 1992).

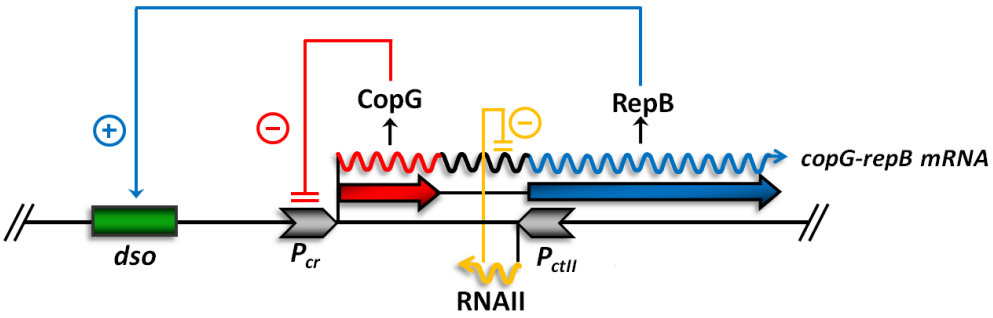


Figure 3. Control-circuit of pMV158 replication. Scheme of the LIC module of plasmid pMV158, displaying the positive action of RepB by binding to the *dso*, as well as the elements of negative control: CopG, which represses transcription from the P_{cr} promoter, and RNAII, which once is paired to its complementary sequence, hinders the binding of the ribosome to the *repB* translation-initiation signals.

The second element, RNAII, participates in control of replication but acting at a post-transcriptional level. RNAII is synthesized from the P_{ctII} promoter and is an antisense-RNA of ~50 nt (del Solar & Espinosa 1992). Binding of this inhibitory RNA to its complementary sequence in the *copG-repB* mRNA does not block directly any element required for efficient *repB* translation, although formation of the RNA-RNA duplex might hinder the binding of the ribosome to the *repB* translation-initiation signals (López-Aguilar, Ruiz-Masó *et al.* 2013). RNAII is a strong *inc* determinant when supplied *in trans* at high dosage. Mutations that abolish the synthesis of this antisense-RNA also lead to an increase in copy number and, although replication is still controlled by CopG, great fluctuations in copy number within individual cells are observed (del Solar, Acebo *et al.* 1995).

3. The RHH proteins

The ribbon–helix–helix (RHH) superfamily of DNA-binding proteins uses a conserved three-dimensional structural motif to bind to DNA in a sequence-specific manner. This superfamily is typified by *E. coli* methionine repressor MetJ protein (Rafferty, Somers *et al.* 1989) and transcriptional repressor Arc of the *Salmonella enterica* serovar Typhimurium bacteriophage P22 (Breg, van Opheusden *et al.* 1990) (Figure 4). Other members include the plasmid-encoded partitioning ParG protein (Golovanov, Barillà *et al.* 2003) and the ω regulator of the Inc18 plasmids family (Weihofen, Cicek *et al.* 2006). Most members of the RHH superfamily are transcriptional repressors, with some exceptions like AlgZ, which is both an activator and a repressor in *Pseudomonas aeruginosa* (Ramsey, Baynham *et al.* 2005).

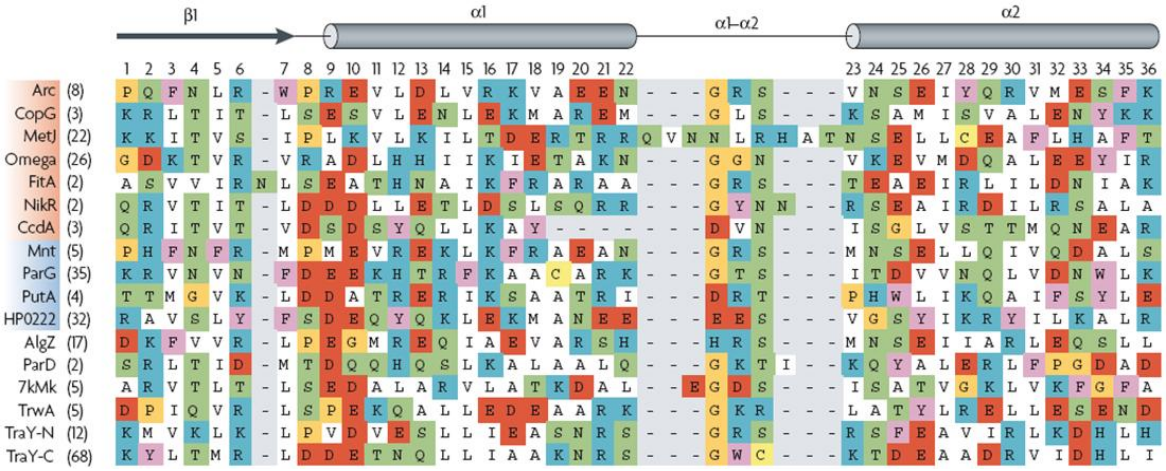
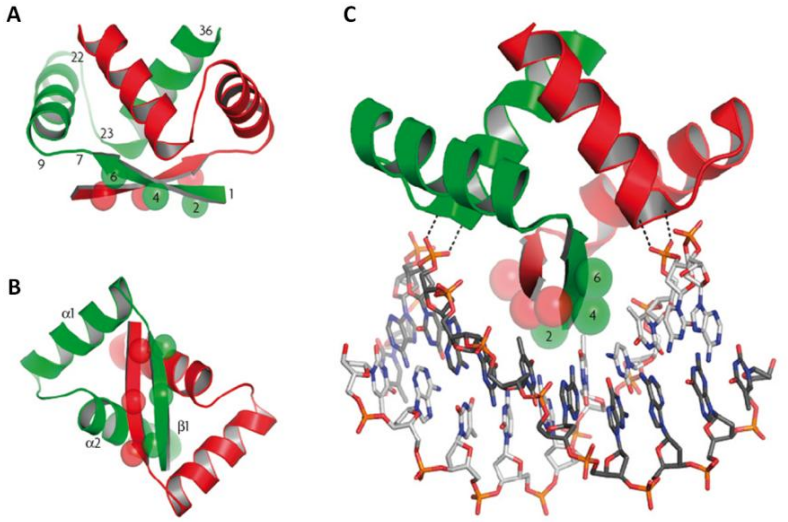


Figure 4. Sequence alignment of characterized RHH proteins. Only the RHH region of each sequence is displayed. Numbers beside protein names indicate the amino acid that is found at motif position one. Numbers above the alignment indicate the Cα positions used for structural alignment. Secondary structure elements are labeled above the alignment. Sequences with names *shaded red* have structures that were determined in complex with operator DNA. Sequences with names *shaded blue* have structures that were determined for the RHH sequence alone. The structures of sequences with no shading have not been determined but are predicted to be RHH proteins based on other experimental evidence. Amino acids are shaded by property: *cyan*, positive charge; *red*, negative charge; *white*, hydrophobic; *green*, neutral hydrophilic; *purple*, aromatic; *orange*, glycine or proline; *yellow*, cysteine. Taken from (Schreiter & Drennan 2007).

The functional unit of these proteins is a dimer, in which two RHH motifs (40-45 residues) are tightly intertwined to form a stable domain (RHH_2) that has a two-fold symmetry and is capable of binding to DNA; the two-stranded antiparallel β-sheet fits snugly into the major groove of the target DNA, with a turn, usually mediated by glycine, connecting helices of each chain (Figures 4 and 5). So far, the exception to this is MetJ, which does not possess this turn, but helices are connected by a longer loop (Figures 4 and 11) (Rafferty, Somers *et al.* 1989). This two-fold symmetric domain is usually formed by homodimerization, but there are exceptions. That is the case of the F plasmid encoded conjugative protein TraY and the *Agrobacterium tumefaciens* T-DNA regulatory protein VirC2, which contain two repeats of the RHH structural motif in a single polypeptide chain (one N-terminal and one C-terminal) (Bowie & Sauer 1990; Lum & Schildbach 1999; Lu, den Dulk-Ras *et al.* 2009).

The N-terminal residues forming the β-strands mediate much of the DNA sequence-specificity. This is a defining characteristic of the RHH superfamily, and distinguishes these proteins from the HTH superfamily, which insert an α-helix into the major groove (Aravind, Anantharaman *et al.* 2005). A conserved set of non-specific contacts to the DNA phosphate backbone are made by the two protein-backbone amide nitrogens at the N-terminus of the α2 helix of the RHH_2 domain, on either side of the major groove. This interaction is electrostatically favorable because the positive dipole at the helix N-terminus is orientated directly towards a negatively

charged phosphate of the backbone. These contacts anchor the RHH_2 domain to the DNA and orientate the base-contacting β-sheet correctly for optimal interaction with the DNA bases.



Schreiter & Drennan (2007) *Nat. Rev. Microbiol.* **5**(9):710-720

Figure 5. The overall structure and DNA binding of RHH proteins. Three views of the RHH dimer (RHH_2) domain are shown as cartoons and colored by subunit. The three amino acid positions (2, 4 and 6) from each subunit that make sequence-specific nucleotide base contacts are shown as semi-transparent spheres. (A) Reference positions within the RHH motif are numbered on the green subunit. The numbers correspond directly to the sequence alignment in Figure 4. (B) The secondary structure elements are labeled on the green subunit, which also corresponds to the alignment in Figure 4. (C) Shows the interaction of a RHH_2 domain with a DNA operator. Nonspecific anchoring contacts between the N-terminus of the α2 helix and the DNA phosphate backbone are shown as *dashed black lines*. Specific base contacts are made by positions 2, 4 and 6 of each subunit from the β-sheet within the major groove of the DNA.

Although the overall DNA binding mode of each individual RHH_2 is the same, the details of the specific DNA base contacts do not seem to be conserved across the RHH superfamily. Each family member binds to a unique operator sequence (or sequences). In general, RHH proteins use one or more amino acid side chains from each N-terminal β-strand (from positions 2, 4 and 6; Figures 4 and 5) to make direct nucleotide base contacts. Although the identities of these three side chains vary across the family (Figure 4), Lys or Arg at position 2 or 6 of the β-strand typically make the largest number of specific contacts. Currently no “code” exists that directly relates the amino acid identity within the β-strand to a specific DNA sequence (Figure 6) (Schreiter & Drennan 2007).

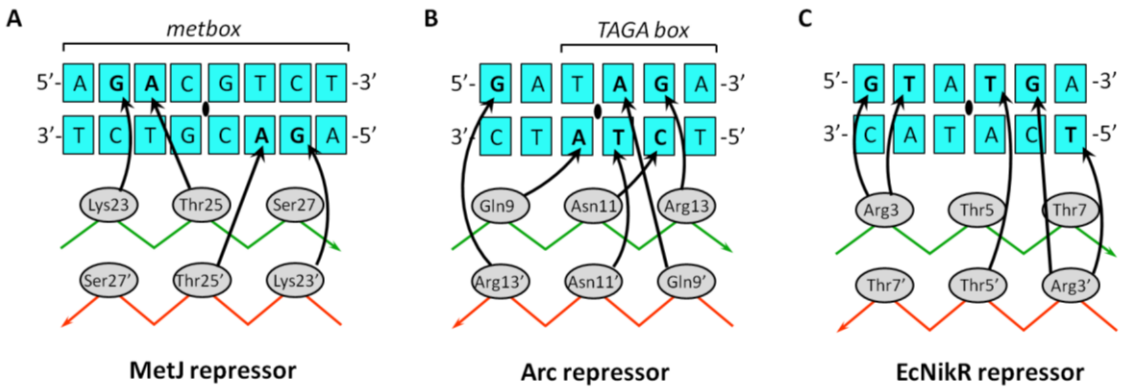


Figure 6. Specific interactions between MetJ, Arc and EcNikR β-strands and DNA operator subsites. Schematic representation of the β-sheet-DNA hydrogen bonding network between the three amino acid side chains of each β-strand of the RHH_2 motif and the nucleotide bases within a DNA operator subsite. Data from structures of the DNA complexes of (A) the MetJ repressor with the 8-bp metbox consensus sequence (Somers & Phillips 1992), (B) the Arc repressor with the right subsite of Arc-operator (Raumann, Rould *et al.* 1994) and (C) the *E. coli* NikR regulatory protein with a subsite of its symmetric operator (Schreiter, Wang *et al.* 2006). Contacted DNA bases are highlighted in *bold*. Hydrogen bonds are indicated by *arrows*. The locations of the dimer pseudo-symmetry axis are indicated by an *ellipse* in the binding sequences. The β-strands from each protomer in the RHH_2 are shown distinctly by *red* and *green lines*.

It is exceedingly rare that protein-DNA recognition is mediated by a single structural unit. Even when only a single protein is involved in DNA binding there are often several independent structural domains that contact the DNA. The RHH motif can be present within the amino acid sequence of a protein either as an isolated RHH motif or as part of larger proteins that have additional domains, which, in turn, can also be directly involved in the DNA-binding process (Figure 7).

Transcriptional regulators commonly bind to adjacent DNA regulatory sites (constituting *operators*) near the promoters they control. Proteins bound at tandem sites of this type have the potential to interact through cooperative protein-protein contacts, and to increase both the affinity and the specificity of DNA binding. The RHH proteins bind as multiple dimeric units to operators that contain repeats of a subsite. In a cooperative interaction, the presence of one protein molecule on a DNA subsite increases the affinity of a second protein molecule for an adjacent binding site (Mao, Carlson *et al.* 1994). Characterizing the contributions of these cooperative interactions is critical to fully understand the process of protein-DNA recognition. To illustrate the general and particular features of the RHH-superfamily proteins, some well-characterized RHH regulation factors are described below.

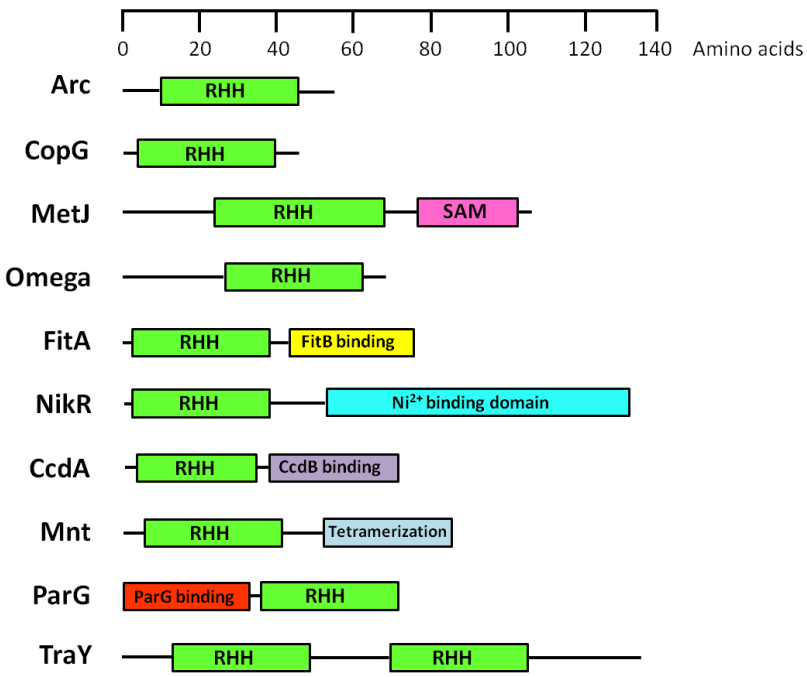


Figure 7. Domain organization of RHH transcription factors. The linear amino acid sequence of each RHH protein is shown as a *black line* on which *colored boxes* and labels that represent functional domains are superimposed. A scale for the amino-acid number is provided at the top. Taken from (Schreiter & Drennan 2007).

3.1. Arc

Arc is a homodimeric repressor of 53 residues encoded by P22, a temperate bacteriophage of *Salmonella typhimurium*. This protein regulates transcription from two overlapping divergent promoters P_{ant} and P_{mnt} in the *immunity I* operon (Figure 8) (Sauer, Krovatin *et al.* 1983). *Immunity I* operon consists of three genes, two of which, *antirepressor* and *arc*, are transcribed together from P_{ant}. The third gene, *mnt,* is transcribed from the P_{mnt} promoter. The main function of this operon is the regulated production of Antirepressor protein. Antirepressor protein antagonizes the P22 c2 repressor during early lytic growth or superinfection. However, *antirepressor* expression is not desired during lysogeny, when its expression would lead to lysis. Thus, two regulators, namely Arc (for antirepressor control) and Mnt (for maintenance of lysogeny), exist to prevent expression of *antirepressor* at different times during the life cycle of the phage. The role of Mnt protein is to block transcription from P_{ant} during lysogeny and to activate its own transcription from P_{mnt} (Vershon, Liao *et al.* 1987a; Vershon, Liao *et al.* 1987b), while the role of Arc is to repress transcription from both P_{ant} and P_{mnt} during late lytic growth (Vershon, Liao *et al.* 1987b).

Arc binds to a 21-bp sequence that has an axis of approximate two-fold rotational symmetry passing through the central base-pair (Vershon, Liao *et al.* 1987b; Vershon, Kelley *et al.* 1989). Thus, the operator can be thought of as being composed of symmetrically related 10-bp half

sites (Figure 9C). Arc is a dimer at high protein concentrations but exists largely as a denatured monomer at the subnanomolar concentrations where the operator DNA binding is observed *in vitro* (Vershon, Youderian *et al.* 1985; Brown, Bowie *et al.* 1990); however, there is no evidence that monomers bind to DNA operator. Under these conditions, operator binding is a highly cooperative fourth-order reaction with a Hill constant of 3.5, and the Arc tetramer is the oligomeric species that is stably bound to operator DNA (Figure 9C) (Brown, Bowie *et al.* 1990). Assembly of the Arc tetramer-operator complex occurs by sequential addition of dimers to operator half-sites. When the left or right operator subsite is occupied by an Arc dimer, cooperative interactions increase the affinity of the second dimer by approximately 5900-fold (Brown & Sauer 1993).

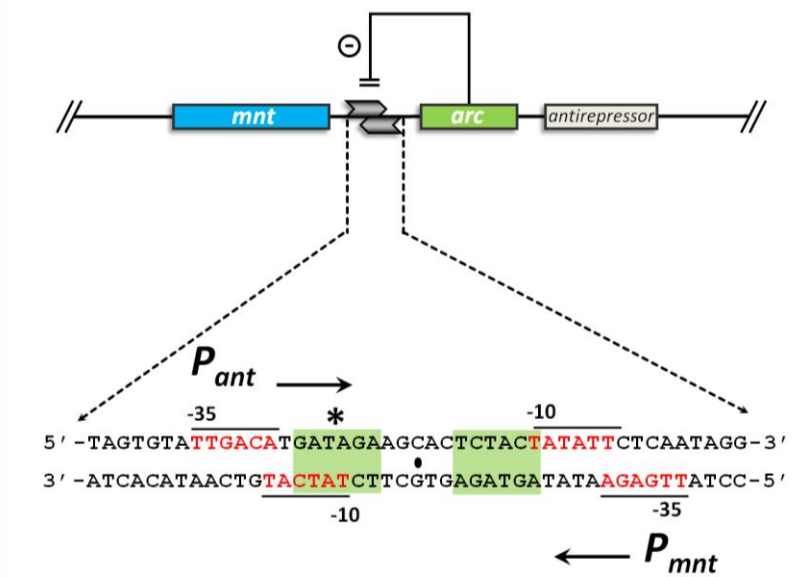


Figure 8. Wild-type *immunity I* region of bacteriophage P22. The -35 and -10 regions of the P_{ant} and P_{mnt} promoters are indicated. The 6-bp sequences to which Arc dimers make base-specific contacts in the Arc-operator co-crystal structure are *green-shaded* (Raumann, Rould *et al.* 1994). When bound, Arc dimers interact to form a cooperatively-stabilized tetramer on the DNA. Subsite marked with an *asterisk* is the primarily responsible for repression of transcription initiation from both promoters (Smith & Sauer 1996).

The DNA sequences of the two operator subsites and their affinities for Arc are similar but not identical (Brown & Sauer 1993). Protein-protein interactions between adjacently bound Arc dimers stabilize DNA binding (Raumann, Rould *et al.* 1994) and Arc variants with normal subsite binding but diminished cooperativity are poorer repressors *in vivo* than wild-type Arc (Smith & Sauer 1995). An Arc dimer bound to a single subsite reduces the rate of RNAP-open complex formation and represses transcription from P_{ant} and P_{mnt} promoter variants to varying degrees (Smith & Sauer 1996). Occupancy of the subsite proximal to the P_{ant} -35 region and the P_{mnt} -10 region (subsite with *asterisk* in Figure 8), results in stronger repression of both promoters than

occupancy of the other subsite. Thus, this operator subsite (left in Figure 8) appears to play the dominant role in mediating repression, and the other subsite (right in Figure 8) may be needed primarily to allow cooperative stabilization of the dimer bound to the left subsite (Smith & Sauer 1996).

Structures of the Arc homodimer have been determined by nuclear magnetic resonance (NMR) (Breg, van Opheusden *et al.* 1990; Bonvin, Vis *et al.* 1994) and by X-ray diffraction (Raumann, Rould *et al.* 1994). The solution and crystal structures are essentially identical. In both cases, monomers wrap around each other to form a roughly globular structure (Figure 9). In the RHH_2 motif, which comprises practically the full-length protein, the two-stranded β-sheet is formed by antiparallel pairing of residues 8-14 of each monomer, and the α-helices are formed by residues 15-30 (α1) and by residues 32-48 (α2) (Figure 9A). The first six N-terminal Arc residues, disordered in the absence of DNA, form tandem reverse turns in the complex that are stabilized by interactions with DNA, with no major differences between the unbound and the Arc-DNA complex structures (Figure 9) (Raumann, Rould *et al.* 1994). In the co-crystal structure, the β-sheet and the N-terminal arms of an Arc dimer lie in the major groove of the DNA, positioned to contact a DNA subsite (Raumann, Rould *et al.* 1994). The surface side chains of the β-sheet form hydrogen bonds with the DNA bases and the N-terminal arms and backbone amides at the N-terminal end of the α2 helix make additional contacts with a number of DNA phosphates. The base recognition is made by Gln9, Asn11, and Arg13 from each strand (Figure 9). Five of these six side chains interact with the first three base pairs of the TAGA box and also hydrogen bond with each other, to form an intricate network at the protein-DNA interface. In addition, conformational adjustments in the β-sheet appear to play important roles in the recognition process. For example, in the unbound protein, the Phe10 and Phe10' side chains are buried in the hydrophobic core, but each side chain swings out and packs between adjacent, unesterified, phosphate oxygens in the complex (Figure 9D) (Schildbach, Karzai *et al.* 1999).

The cooperative contacts between the two Arc dimers bound occur above the center of the operator site and involve only a few residues at the junction of α1- and α2-helices. The key interactions are made by the side chains of the two Arg31’ residues, each of which reaches across the tetramer interface to make two hydrogen bonds with the backbone carbonyl of Asn29’ in the other subunit. Biochemical and genetic studies have shown that changing the spacing between operator half-sites by even one base-pair drastically reduces Arc binding (Vershon, Kelley *et al.* 1989; Brown & Sauer 1993). Since each dimer is relatively rigid, the interaction surface is highly sensitive to even small changes in distance or orientation. As a

result, the formation of a stably-bound tetrameric complex requires the precise spacing of half-sites observed in the wild-type operator.

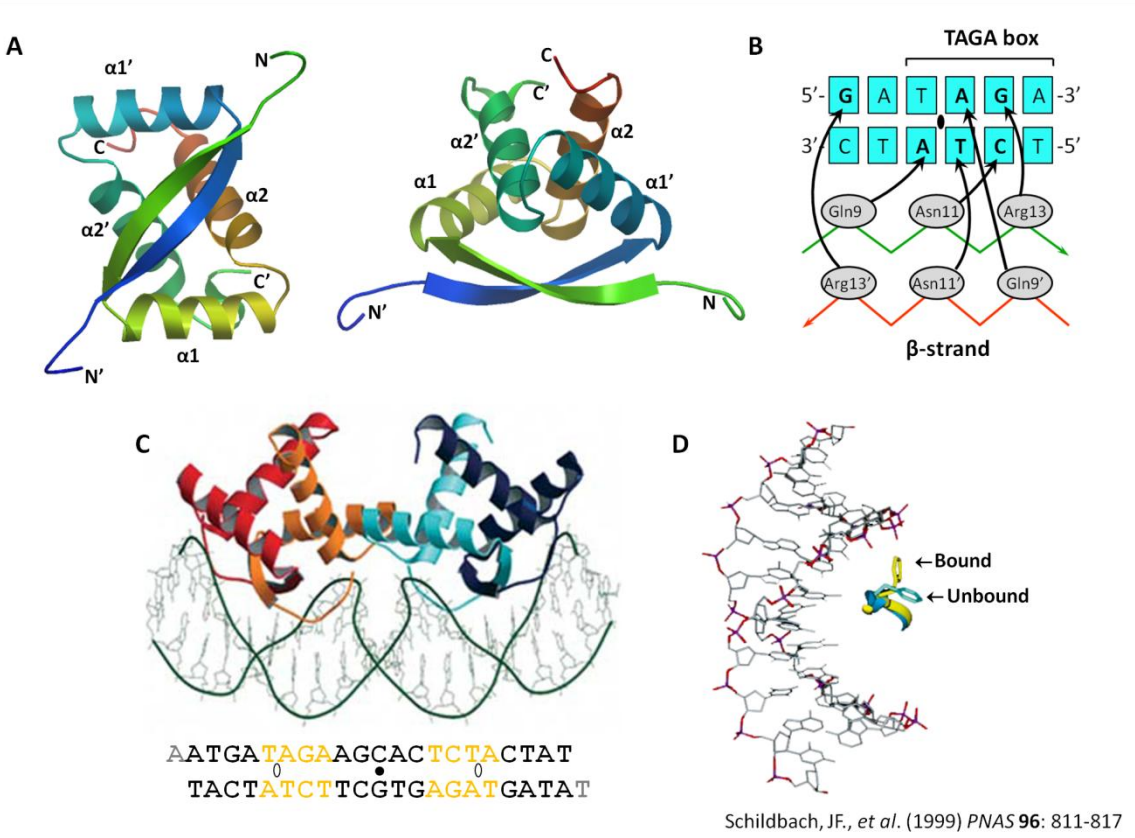


Figure 9. Structures of Arc repressor alone and bound to its operator DNA. (A) Ribbon plot of two different views of the Arc RHH_2 motif. The helices and chain termini are labeled. Symmetric elements in different protomers are marked by *primes*. Taken from PBD (ID: 1ARR). (B) Base interactions of the β-sheet with the nucleotides in the left subsite of Arc operator (Raumann, Rould *et al.* 1994). (C) Ribbon plot of the Arc tetramer-DNA complex. Subunits of a dimer are indicated by variations of color intensity. The N-terminal arms make contacts with DNA backbone and alter the idealized B-DNA conformation. Taken from (Spector, Sauer *et al.* 2004) PDB (ID: 1PAR). In *low panel*, the dsDNA used for co-crystallization, which includes the Arc operator 21-bp sequence. Dyad-symmetric operator half-sites are highlighted in *orange*. The locations of the tetramer and dimer pseudo-symmetry axis are indicated by a *filled circle* and *open ovals*, respectively. (D) Superimposition of the conformation of the Phe10 side chain in unbound (*blue*) and operator-bound Arc (*yellow*).

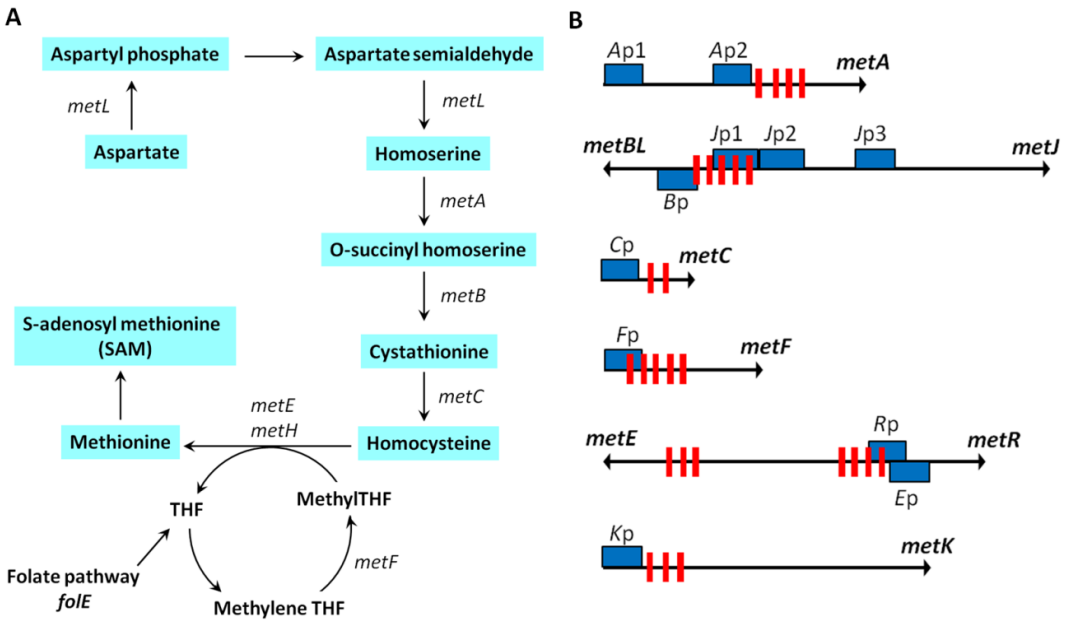
Binding of the Arc tetramer to the operator induces DNA bending. Repressor binding provokes the compression of the DNA minor groove in the center of the operator and, by interaction of the N-terminal arms of Arc, the widening of the major grooves near the center of each subsite (giving an overall bend of about 22°) (Figure 9C) (Raumann, Rould *et al.* 1994; Gomis-Rüth, Solá *et al.* 1998a). The DNA distortions that lead to bending of the DNA appear to be required for tetramer formation, as modeling the dimers on linear DNA indicates that they could not interact cooperatively. N-terminal arm mutants of Arc exhibit significant DNA binding defects, although they still display wild-type hydroxyl radical footprints (Raumann, Rould *et al.* 1994).

3.2. MetJ

The *E. coli* methionine repressor MetJ, a homodimeric protein of 12 kDa, was the first structurally characterized member of the RHH class of DNA-binding proteins (Rafferty, Somers *et al.* 1989). Methionine is a key amino acid in protein biosynthesis, and is also a component of the activated methyl cycle leading to the synthesis of S-adenosyl methionine, a major intracellular methylation agent (Figure 10A). In *E. coli*, the Met regulon consists of ten biosynthetic genes (*metA, merB, merC, metE, metF, metH, metK, metL, metQ* and *metX*), two regulatory genes (*metJ* and *metR*), and the methionyl tRNA synthetase gene (*metG*). Most of these genes do not form operons and are scattered throughout the chromosome (Bachmann 1990; Old, Saint-Girons *et al.* 1993). MetJ represses transcription of at least nine genes encoding proteins involved in the methionine biosynthesis (Figure 10) (Somers & Phillips 1992; Somers, Rafferty *et al.* 1994), including its own gene, *metJ* (Saint-Girons, Duchange *et al.* 1984).

MetJ dimers bind to operator sequences that contain two to five sequential tandem repeats of an 8-bp motif, termed the "metbox" (Figures 10C and 11B) (Phillips, Manfield *et al.* 1989). There is a great deal of sequence variability among metboxes within each operator, but the shared consensus sequence is palindromic 5'-AGACGTCT-3' (Figure 10C) (Phillips, Manfield *et al.* 1989), and *in vitro* selection has shown that this sequence is optimal for MetJ binding (He, Stockley *et al.* 1996). The number of contiguous metboxes as well as the match of those sequences to the consensus are likely responsible for the differential levels of repression of the genes in the Met regulon (Old, Phillips *et al.* 1991; Marincs, Manfield *et al.* 2006). A minimum of two tandem metboxes is required for efficient MetJ binding *in vitro* and repression of transcription *in vivo* (Phillips, Manfield *et al.* 1989). Each metbox-bound repressor dimer makes protein-protein contacts to neighboring dimers making assembly of higher order repressor complexes a cooperative process (Rafferty, Somers *et al.* 1989; Hyre & Spicer 1995).

MetJ repressor is activated by its co-repressor S-adenosyl methionine (SAM). Two positively charged SAM molecules bind to a site on the opposite face of the protein from the DNA-binding motif (Figure 11) (Somers & Phillips 1992). Bound SAM creates an unusual long-range electrostatic interaction with the phosphodiester backbone of DNA and increases the affinity of MetJ for the operator DNA at least 1000-fold (Phillips & Phillips 1994; Parsons, Persson *et al.* 1995). Under *in vitro* conditions, MetJ can bind DNA in the absence of SAM with a much lower affinity (Davidson & Girons 1989), but this situation is unlikely to occur *in vivo* because SAM is an essential metabolite that is required for cell growth and can never be completely depleted in the cell (Wei & Newman 2002).



C

met operator	Operator sequences
metA	**AG**CTA**TCT** ·G**GA**T**GTCT** ·**AA**A**CGTAT** ·**A**AG**CGTAT**
metB	**ATACG**CAA ·**AGA**AG**TTT** ·**AGATGTCC** ·**AGATGTAT** ·**TGACGTCC**
metC	**AGACATCC** ·**AGACGTAT**
metF	**CTTCATCT** ·**TTACATCT** ·**GGACGTCT** ·**AA**A**CG**GA**T** ·**AGATGT**GC
metE	G**GATGAAT** ·**AA**A**CTT**GC ·C**GCCTTCC**
metR	**AGGATTTT** ·**AGCCGTCC** ·**AGATGTTT** ·**A**C**AC**A**TCC**
	Consensus sequence: **AGACGTCT**

Figure 10. Representative components of the *E. coli* methionine biosynthesis and *met* genes operator sequences for MetJ. (A) Summary of the biosynthetic pathway for methionine showing the gene product operating at each step (Marincs, Manfield *et al.* 2006). (B) Architecture of well-known *met* promoters of *E. coli* (Marincs, Manfield *et al.* 2006). The positions of promoters for the labeled genes are indicated by *blue boxes*. *Arrowheads* show the positions of the start codons and the direction of transcription. *Red bars* indicate the operator metboxes, most of which are detailed in *C*. (C) Naturally occurring *E. coli* operator sequences for the methionine repressor MetJ. *Bold* bases denote identity with the consensus sequence, metbox (Somers, Rafferty *et al.* 1994).

The crystal structures of the methionine aporepressor (MetJ) and holorepressor (MetJ-SAM) have been solved (Rafferty, Somers *et al.* 1989). MetJ is a highly intertwined symmetric dimer of 104 residues per subunit (Figure 11A). Each subunit is constituted by one β-strand (residues 21-28) and three α-helices, α1 (residues 30-45), α2 (residues 53-66) and α3 (residues 86-94). The first N-terminal 11 residues are in extended conformation before entering a flexible β-hairpin loop (residues 12-20), which is disordered in some crystal forms and can adopt different conformations (Figure 11). The RHH_2 motif is formed by pairing of the two-stranded antiparallel β-sheet and α1- and α2-helices. Unlike other RHH proteins, α1- and α2-helices are connected by a longer loop instead of a glycine-mediated turn. The α3 helix connects with α2 helix via the loop formed by residues 77-84 and it is immediately followed by a sharp turn at Gly96 before

entering a smaller turn and an extended chain towards the C-terminus at residue 104 (Somers, Rafferty *et al.* 1994). The two SAM molecules bind to symmetrically related sites in the holorepressor, with their purine rings inserted into hydrophobic pockets on the protein surface that are otherwise occupied by the side chains of Phe65 in the aporepressor (Figure 11). Binding of SAM does not induce significant structural changes (Somers & Phillips 1992).

The crystal structures of two repressors bound to a 19-bp oligonucleotide, the sequence of which encompasses the minimal consensus operator sequence, has also been determined (Figure 11D) (Somers & Phillips 1992). Each dimer inserts the β-sheet into the major groove, and makes direct contacts with four of the 8-bp of the metbox via the side chains of Lys23 and Thr25 form each β-strand (Figure 11C). Numerous contacts are made to the sugar-phosphate backbone, positioning the repressor accurately onto the operator sequence. In particular, strong interactions are made by the main chain amide nitrogen atoms of Asn53 and Ser54, and by the side chain of Ser54 (at the N-terminus of α2 helix from each subunit) with the non-esterified phosphate oxygens 5' to the first and last guanines of each metbox (Figure 11E). These interactions are energetically very favorable owing to the positive dipole of the N-terminus of the α-helices (Hol, van Dujinen *et al.* 1978), which point directly at the negatively charged phosphates of the DNA backbone. The two central phosphates that are contacted by Asn53 and Ser54 lie near the junction of the metboxes (Figure 11D, low panel), and are displaced by 2 Å from their expected positions in idealized B-form DNA in a direction that narrows the minor groove. This latter deformation is necessary for the correct positioning of adjacently-bound repressors, each on different faces of the double helix, in order to form the cooperative interactions essential for formation of the repression complex (Figure 11) (Garvie & Phillips 2000).

Higher-order complexes containing more than two metboxes have not been successfully crystallized. However, results of small angle neutron scattering (SANS) with contrast variation measurements of three MetJ complexes with SAM and DNA containing two, three and five metboxes in solution, suggest that the solution conformation of the two-metbox complex differs significantly from the crystal structure (Augustus, Reardon *et al.* 2006). Furthermore, the higher-order complexes, particularly the five-metbox complex, differs significantly from the smaller complexes, providing much closer packing of the adjacent MetJ dimers and allowing additional contacts not available in the crystal structure. This suggests that there is a structural basis for the differences observed in the regulatory effectiveness of MetJ for the various genes of the Met regulon (Augustus, Reardon *et al.* 2006).

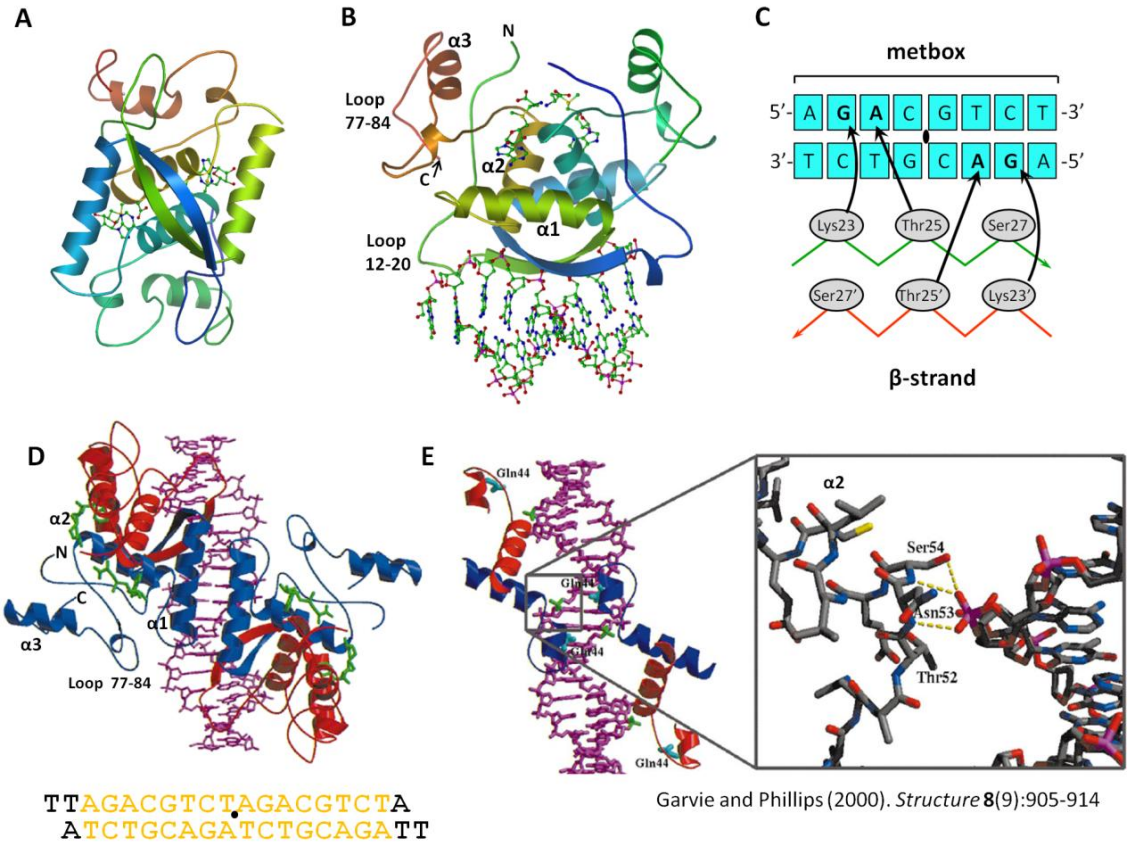
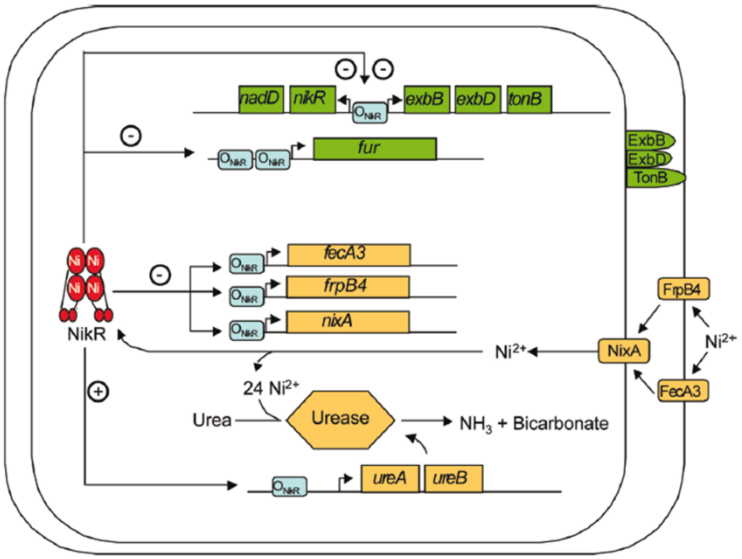


Figure 11. Structure of the holorepressor MetJ and base recognition in the metbox. (A) Ribbon plot of MetJ repressor-corepressor complex, viewed along the molecular dyads. SAM molecules are shown as *balls* and *sticks***.** This is the bottom view of repressor diagram in *B***.** Taken from PBD (ID: 1CMB). (B) Ribbon plot of one MetJ repressor complexed to the consensus operator sequence (*lower,* as *balls* and *sticks*), showing the opposite location of SAM ligands (upper, as *balls* and *sticks*). The β-sheet lies on the major groove of the DNA. Helices and chain termini of a subunit are labeled. Taken from PBD (ID: 1CMA). (C) Main interactions between residues of repressor β-sheet and nucleotides of the metbox. (D) Two repressors (each comprised of a *red* and a *blue* subunit) bind to adjacent 8-bp metboxes (*low panel*), one on the *upper left* and the other at the *lower right*. The repressors make cooperative interactions at the center of the complex via the α1-helices of the *blue* subunits. The SAM molecules and DNA are shown in *green* and *purple,* respectively. *Low panel*, 19-bp oligonucleotide used for co-crystallization by (Somers & Phillips 1992). *Black circle* indicates the symmetry dyad between metboxes. The helices and chain termini of the *blue* subunit of one repressor are labeled. (E) Details of the interaction between the main chain amide nitrogens of Asn53 and Ser54 in the α2 helix with the sugar-phosphate backbone. The ribbon diagram shows the contacted phosphates in *green*; for clarity, only α2-helices and part of the α1-helices are shown. Gln44 from each subunit is shown in *cyan* and is labeled. A *close-up view* of one of the interactions is shown in *stick* representation on the right.

3.3. NikR

Metal ions are essential nutrients for all cells, but their intracellular concentrations and distribution must be tightly regulated to avoid toxicity. NikR is the only known metal-responsive member of the RHH family of transcription factors and functions as a homotetramer (Chivers & Sauer 1999) only in the presence of specific divalent transition metal ions such as Ni^{2+}. Nickel is an essential cofactor in several metalloenzymes (Maroney 1999; Mulrooney & Hausinger 2003) that are expressed primarily in microorganisms, such as the model *E. coli* and the human gastric pathogen *Helicobacter pylori*; and it is most often utilized under anaerobic conditions (Hausinger 1987; Maroney 1999). In these organisms, incorporation of nickel into enzymes is necessary for metabolic adaptation to changing environmental conditions (Mulrooney & Hausinger 2003).

H. pylori utilizes urease, a nickel-cofactor metalloenzyme that converts urea to ammonia and bicarbonate to neutralize the acid conditions of the gastric epithelium (Mobley, Island *et al.* 1995; Mulrooney & Hausinger 2003; Stingl & De Reuse 2005). Although nickel is required for *H. pylori* survival, it can be toxic at high concentrations (Krishnaswamy & Wilson 2000). As a result, *H. pylori* must tightly control intracellular nickel levels. A key protein involved in nickel ion metabolism and homeostasis for *H. pylori* is the nickel regulatory protein, NikR (HpNikR). HpNikR regulates multiple genes as an activator or repressor by binding to varied operator sequences (Dosanjh, West *et al.* 2009). Interestingly, these genes include both genes that encode proteins that utilize nickel to function, such as NixA (a nickel importer), HPN (a nickel storage protein) and UreA (urease), and genes that encode proteins that do not directly use nickel, such as genes involved in iron uptake and storage (e.g., *pfr* and *fur*), genes involved in motility (e.g., *cheV*, *flaA*, and *flaB*), genes involved in stress response (e.g., *hrcA, grpE* and *dnaK*) and genes that encode outer membrane proteins (e.g., *omp11, omp31*, and *omp32*). Probably because of the comparatively small size of the *H. pylori* genome, HpNikR has acquired multiple functions (van Vliet, Ernst *et al.* 2004). Figure 12 pictures an overview of the regulation exerted by HpNikR on several of these genes.



Dosanjh NS., *et al.* (2009) *Biochemistry*, **48**(3): 527-536

Figure 12. Scheme of the regulation exerted by HpNikR on several genes. HpNikR activates *ureA* transcription and represses *nixA*, *exbB*, *fecA3* and *frpB4* transcription, and its regulatory role for *fur* has not been defined.

The recognition sequences in each promoter are defined by a series of poorly conserved 6-bp imperfect inverted repeat half-sites separated by a 15-bp spacer (Delany, Leva *et al.* 2005; Benanti & Chivers 2007; Dosanjh, West *et al.* 2009). HpNikR binds to these promoters with a range of affinities (Benanti & Chivers 2007; Dosanjh, West *et al.* 2009), utilizes a two-tiered mode of DNA recognition by binding to some gene promoters with high affinity and others with low affinity. The genes that are tightly regulated by HpNikR encode proteins that utilize nickel, while those that are less tightly regulated encode other types of proteins. The promiscuity of DNA binding by HpNikR contrasts with high sequence selectivity of the NikR proteins from *E. coli* (EcNikR) (Chivers & Sauer 2000) and *Geobacter uraniireducens* (GuNikR) (Benanti & Chivers 2010), which recognize highly conserved inverted repeats found in only one or two promoter regions per genome.

The 133-amino acid orthologue from *E. coli* has been characterized most extensively at the molecular level. In *E. coli*, nickel is acquired via an ABC-type membrane transporter, NikABCDE (Navarro, Wu *et al.* 1993). This transporter is synthesized under anaerobic conditions to meet the increased demand for nickel resulting from hydrogenase synthesis (Wu, Mandrand-Berthelot *et al.* 1989; Wu, Navarro *et al.* 1991; Navarro, Wu *et al.* 1993). Regulation of *nikABCDE* expression is positively controlled by FNR, a fumarate and nitrate reduction regulator (Wu, Mandrand-Berthelot *et al.* 1989), and negatively controlled by EcNikR (Chivers & Sauer 2000), in both cases by direct protein binding to the *nikABCDE* promoter ($P_{nikABCDE}$ or P_{nik}). This arrangement provides two distinct inputs that control nickel uptake. A decrease in oxygen tension results in activation of FNR and upregulation of *nikABCDE* expression, while the presence of excess nickel activates EcNikR, which overrides the action of FNR and results in repression of

nikABCDE transcription. Repression of transcription of the *nik* operon in presence of nickel is the only one function ascribed for EcNikR (De Pina, Desjardin *et al.* 1999; Chivers & Sauer 2000; Rowe, Starnes *et al.* 2005). EcNikR therefore serves as a cytoplasmic nickel sensor, stopping production of the nickel importer when intracellular levels of nickel are sufficient.

EcNikR recognizes a 28-bp palindromic operator sequence (GTATGA-N_{16}-TCATAC) within the P_{nik} (Chivers & Sauer 2002). In the apo-crystal structure, EcNikR is a homotetrameric protein composed of two types of domains: a central tetrameric metal binding domain (MBD) and two flanking dimeric RHH domains (Figure 13A) (Chivers & Sauer 2002; Schreiter, Sintchak *et al.* 2003; Schreiter, Wang *et al.* 2006). This spatial separation of the two DNA-binding domains explains how EcNikR can recognize two operator subsites distantly separated by two turns of DNA. However, this structure is not suitable for interaction with operator DNA (Schreiter, Sintchak *et al.* 2003).

EcNikR has multiple types of metal binding sites: high-affinity nickel binding sites for stoichiometric nickel (low picomolar dissociation constant) (Chivers & Sauer 2002; Wang, Dias *et al.* 2004), potassium binding sites (Schreiter, Wang *et al.* 2006; Phillips, Nerenberg *et al.* 2009; Wang, Li *et al.* 2010), and low-affinity nickel binding sites (dissociation constants from 30 µM to 30 nM) (Chivers & Sauer 2002; Bloom & Zamble 2004). The four identical high affinity nickel binding sites are located at the central tetrameric interface in the MBD (Figure 13A) (Carrington, Chivers *et al.* 2003; Schreiter, Sintchak *et al.* 2003; Schreiter, Wang *et al.* 2006). Nickel ions bind with a square planar-coordination geometry, ligated by the protein side chains His87, His89, and Cys95 from one EcNikR subunit, as well as the side chain of His76' of an adjacent subunit (Figure 13A, right). Although binding of nickel to the high-affinity site of EcNikR does not directly alter the conformation of the RHH domains for DNA binding, it does have a very important short-range effect on the conformation of the protein. The RHH domains of NikR are connected to the MBD by flexible linkers that vary significantly in length and amino acid sequence as a function of species. Residues 62-80 of EcNikR in this flexible linker are disordered in the absence of nickel but form a loop and stable α-helix (α3) when nickel is bound because of the direct involvement of His76' in nickel coordination (Figure 13B) (Schreiter, Sintchak *et al.* 2003). The EcNikR-DNA complex structure shows the importance of this conformation change in creating 6-8 polar interactions between the loop containing residues Lys64 and Arg65 and the phosphate backbone of the operator DNA, thereby initiating EcNikR-DNA binding (Figure 13B) (Schreiter, Wang *et al.* 2006; Phillips, Schreiter *et al.* 2008). Complementary interaction of the nickel-stabilized MBD with DNA localizes EcNikR to the DNA helix and allows cooperative interactions

of both RHH domains. Stable binding to the operator then occurs when the specific operator base sequence is recognized and additional specific base contacts can be made.

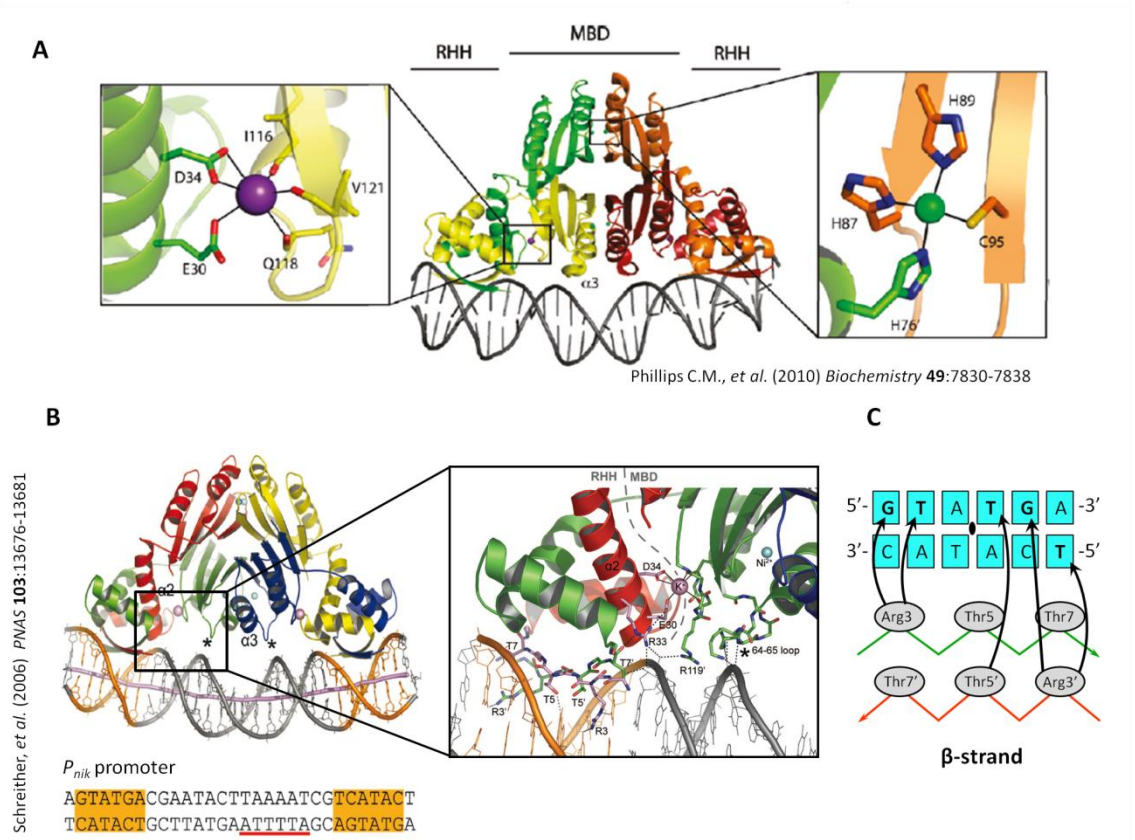


Figure 13. The metal binding sites of EcNikR and the structure of the EcNikR–operator DNA complex. (A) Overall topology of EcNikR with the metal binding domains (MBD) and ribbon-helix-helix (RHH) domains indicated. Representative structures of the potassium (left panel) and high-affinity nickel sites (right panel) with coordinating amino acids labeled. (B) The EcNikR–DNA complex with the NikR tetramer colored by subunit, and DNA displayed as *sticks* with a cartoon tube tracing the phosphorus positions. The dyad-symmetric operator half-sites are *colored orange*, and the DNA helical axis is shown as a *purple tube*. Nickel and potassium ions are shown as *cyan* and *pink spheres*, respectively. *Asterisks* indicate the MBD loop that contacts DNA. The α2 helix, which contributes the conserved secondary metal site ligands E30 and D34, is labeled on the red NikR molecule. The α3 helix, which is structurally stabilized upon high-affinity nickel binding, is labeled on the blue NikR molecule. *Close up*: a *dashed grey line* delineates the RHH and MBD. Bonds to the potassium ion at the domain interface are shown as *solid black lines*, and hydrogen bonds as *dashed black lines*. Portions of NikR contacting DNA or the secondary metal site are shown as *sticks*, and important side chains are labeled. An *asterisk* indicates the MBD loop, containing Lys64 and Arg65, which contacts DNA. *In low panel*, the dsDNA used for EcNikR co-crystallization, which includes the wild-type EcNikR operator sequence. Dyad-symmetric operator half-sites are *orange-shaded*, and the -10 region of the P_{nik} promoter is underlined with *red*. (C) Hydrogen bonds established by the β-sheet residues with nucleotides in a half-site operator.

In the DNA-bound EcNikR tetramer, two identical octahedral metal binding sites are formed between the MBD and RHH domains (Figure 13A) (Schreiter, Wang *et al.* 2006). Potassium binding to this site helps to stabilize the RHH domains in DNA-binding conformation, increasing the affinity of protein for DNA (Phillips, Nerenberg *et al.* 2009). The strongest binding of EcNikR to operator DNA (apparent affinity 10-15 pM) is observed in the presence of nickel

concentrations that are more than 10^6-fold higher than those required to bind the high-affinity binding sites (Chivers & Sauer 2000). Crystallographic data suggest that there are a number of low-affinity nickel sites on EcNikR (Phillips, Schreiter *et al.* 2010). These sites are all on the surface of the protein, and five of the six site types have at least one histidine ligand. The binding of excess nickel ions to the low-affinity sites does not induce significant conformational changes in protein, but enhances protein affinity for DNA *in vitro* primarily through an electrostatic effect (Phillips, Schreiter *et al.* 2010). All low-affinity sites identified are easily accessible to excess nickel ions, even if EcNikR is already bound to DNA. In this sense, the binding of low-affinity nickel to the surface of EcNikR could help tune the affinity of the protein for the *nik* operon in response to the nickel concentration in the cell, either by facilitating the EcNikR-DNA binding interaction or by stabilizing the bound EcNikR-DNA complex (Figure 14) (Phillips, Schreiter *et al.* 2010).

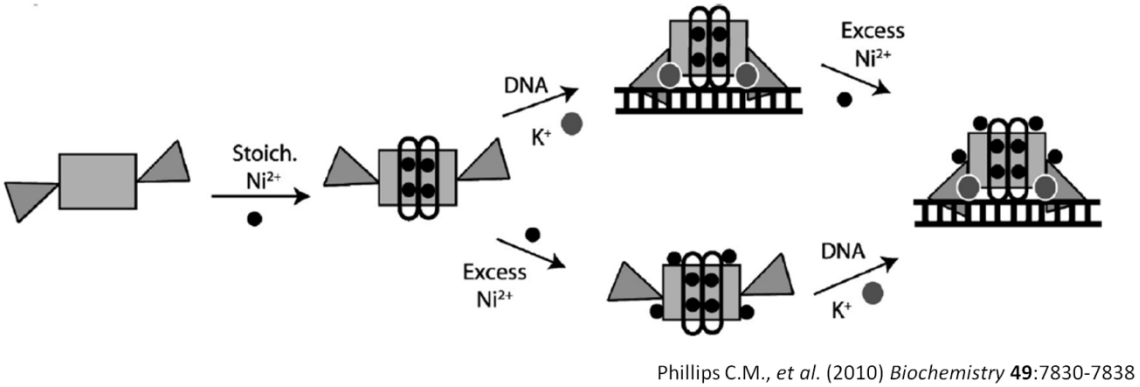


Figure 14. Proposed scheme for EcNikR binding to DNA considering all three types of metal binding sites. Nickel ions (*black circles*), potassium ions (*gray circles*), DNA (*ladder*), RHH domains (*triangles*), ordered α3-helices (*white ovals*), and the central MBD (*rectangle*) are illustrated. As structural data indicate that excess nickel ions can bind to the protein before and after DNA binding, two pathways are depicted.

4. Transcriptional repressor CopG

The pMV158-encoded CopG protein represents the smallest natural transcriptional repressor characterized so far, and is also the prototype of a series of repressor proteins encoded by RCR plasmids that exhibit a similar genetic structure at their leading-strand initiation and control regions (del Solar, Giraldo *et al.* 1998; Gomis-Rüth, Solá *et al.* 1998a; del Solar, Hernández-Arriaga *et al.* 2002). A sequence alignment of CopG with other Cop proteins of the pMV158 family of plasmids shows that they all display features that are compatible with the structure described for CopG as a RHH protein (Figure 15) (Gomis-Rüth, Solá *et al.* 1998a; del Solar, Hernández-Arriaga *et al.* 2002). Varying lengths of the members are accounted for by longer N- and C-terminal segments. All sequences display the glycine-mediated turn connecting two α-

helices and similar residues at key positions, i.e. those required for the maintenance of the hydrophobic core structure of the dimer. Thus, CopG-like proteins of replicons belonging to the pMV158 family are likely to share the RHH motif.

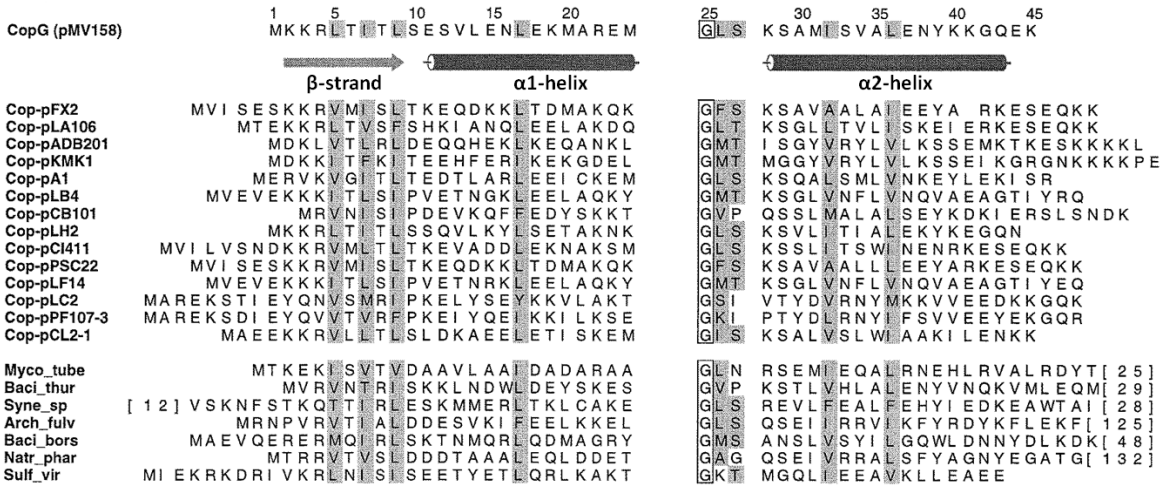


Figure 15. Sequence alignment of plasmid-encoded CopG-like proteins. Some Cop proteins encoded by RCR plasmids of the pMV158 family were aligned with the CopG sequence. The glycine residues of the turn connecting α1- and α2- helices are framed. Conserved (hydrophobic) positions are highlighted by *grey background*, as are the highly conserved Thr/Ser residues in the mentioned turn. The numbering and regular secondary-structure regions correspond to the CopG structure. Taken from (Gomis-Rüth, Solá *et al.* 1998a).

CopG is a homodimeric protein of 45 residues per subunit. The polypeptide chain of CopG starts with a segment in extended conformation reaching Leu9 (β-strand) (Figure 16A). At Glu11, the α1 helix starts with its axis rotated ~70° away from the direction of the β-strand. This helix finishes with Met24, and then enters a glycine-mediated turn leading to α2 helix, which begins with Lys28 and extends to the C-terminal end. The axis of the α2 helix is rotated ~90° with respect to α1 helix, because of the presence of Gly25 (Figure 16A). The latter bihelical topology is the same as that found in proteins with an HTH (Helix-Turn-Helix) motif involved in DNA binding (Matthews, Ohlendorf *et al.* 1982; Pabo & Sauer 1984); however, in CopG this element has a role in oligomerization instead of in DNA recognition (Gomis-Rüth, Solá *et al.* 1998a). As being a RHH protein, the association of protein as a dimer leads to the formation of an antiparallel β-ribbon, which is presented to the exterior from a hydrophobic and dimeric protein core, mainly formed by internal residues coming from both strands and the four helices.

The hydrodynamic behavior of the native protein shows that CopG is a dimer in solution at concentrations ranging from 10 to 800 μM, with no detectable monomers or association states with a higher number of molecules than dimers. The CopG dimers are nearly spherical, with a

deduced Stokes radius of around 16 Å (Acebo, García de Lacoba *et al.* 1998), which permits the CopG dimers to occupy approximately one helical turn when bound to its target DNA.

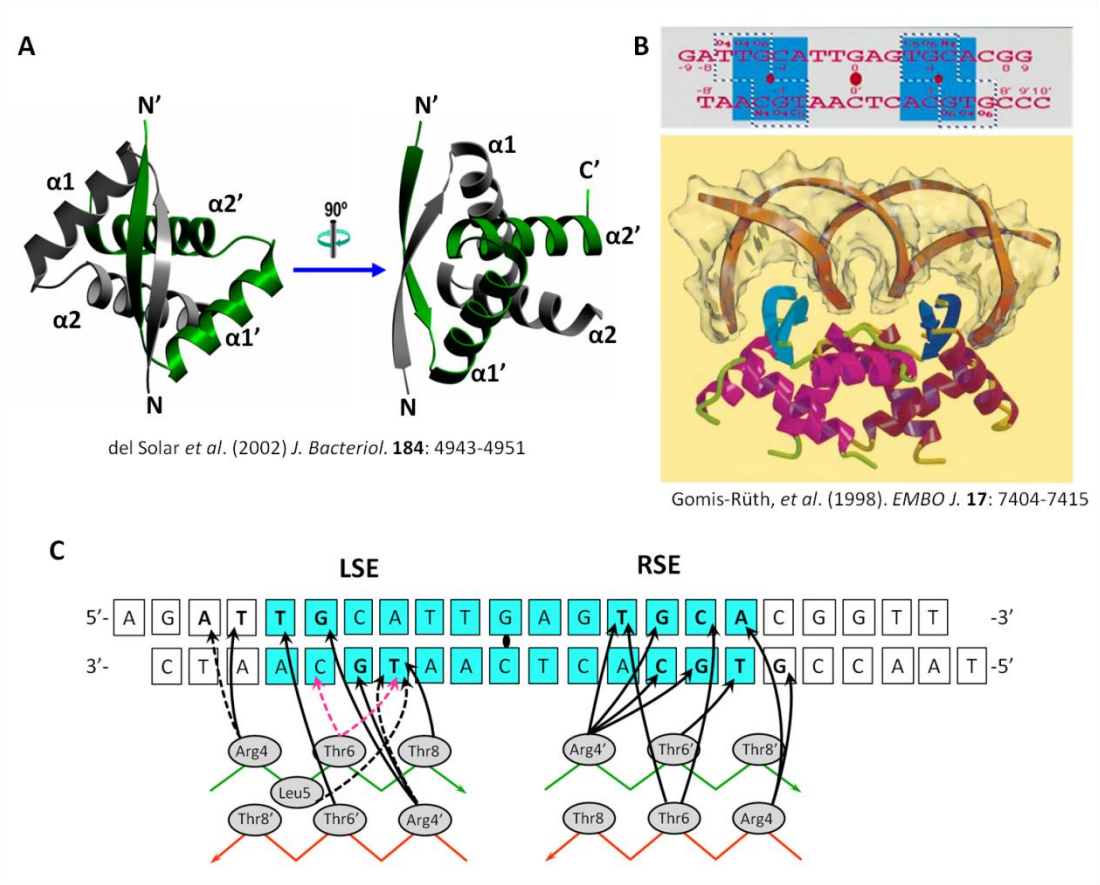


Figure 16. Structure of CopG and its interactions with DNA in the co-crystals. (A) Three-dimensional structure of CopG dimer. The helices and chain termini of the subunits are labeled. Symmetric elements in different protomers are marked by *primes*. (B) The structure of two CopG dimers complexed with the 19-bp dsDNA containing the SE, superimposed with its semi-transparent Connolly surface. In *upper panel*, sequence of the 19-bp dsDNA used for co-crystallization. The two-fold symmetry axis of the SE is denoted by a *large ellipsoid* at the center of the sequence. Two symmetrical sequences are indicated by *blue boxes* left and right of the center. Their internal two-fold symmetry is denoted by the *small ovals.* The bases recognized by CopG are indicated by *dashed boxes*. Base atoms participating in DNA-protein interactions are indicated. (C) Scheme of the specific contacts (*arrows*) observed in the co-crystals of CopG bound to a 22-bp dsDNA comprising the SE (*blue shaded bases*). Contacted DNA bases are highlighted in *bold*. *Dashed black arrows* indicate the interactions not observed in the 19-bp dsDNA co-crystals and *pink dashed ones* indicate those observed in the 19-bp dsDNA co-crystals that were not detected in the co-crystals with the 22-bp dsDNA. The *black dot* marks the symmetry axis of the 13-bp SE. The β-strands from each protomer are shown distinctly by *red* and *green lines*.

DNAse I protection assays indicate that the specific target of CopG spans about 50 bp and includes a central 13-bp (pseudo)symmetric element (SE) that comprises most of the -35 region of the promoter (Figure 17A). Each half-site of the SE, namely Left and Right Symmetric Elements (LSE and RSE, respectively), contains the palindromic sequence 5'-TGCA-3', so that two self-

symmetric subsites are present within the operator (Figure 16B) (del Solar, Pérez-Martín *et al.* 1990; Gomis-Rüth, Solá *et al.* 1998a).

Three-dimensional structures of CopG bound to either a 19-bp (Figure 16B) (Gomis-Rüth, Solá *et al.* 1998a) or a 22-bp (Figure 16C) (Costa, Solà *et al.* 2001) oligonucleotide containing the SE, revealed that two CopG dimers bind to two successive major grooves of the DNA, which, in turns, correspond to the subsites within the SE (Figure 16B and C). Resolution of these structures constituted the first examples of a solved plasmid-encoded transcriptional repressor crystal structure.

No significant differences were observed in the main chains when compared with the free protein structure. CopG interacts with DNA bases via the N-terminal β-ribbon, and with the backbone phosphate groups via residues at the N-terminus of the α2 helix. The SE bound-CopG tetramer induces a bend of 60° in the nucleic acid moiety. This bend is produced by marked compression of both minor and major grooves facing the protein (Figure 16B) (Gomis-Rüth, Solá *et al.* 1998a). Although each CopG dimer has local two-fold symmetry, the two β-strands contact different bases, not matching the self-symmetry (5'-TGCA-3') of the DNA sequence (Figure 16). The base-recognition in the LSE and RSE subsites is asymmetric in relation to the center of the operator, there are more protein-base contacts on the right side (nine) than on the left side (eight) (Figure 16C). The amino acids involved in these interactions are also asymmetric (Costa, Solà *et al.* 2001).

Electrophoretic mobility shift assays and previous data from DNase I and hydroxyl radical footprinting assays strongly suggested that other two dimers would bind laterally to the SE, in the regions located one helical turn apart, defined as the LA (Left Arm) and RA (Right Arm) (Figure 17A) (del Solar, Pérez-Martín *et al.* 1990; Gomis-Rüth, Solá *et al.* 1998a). Therefore, we work with a nucleoprotein model of P_{cr} promoter repressosome that involves the binding of four CopG dimers on the same face of the operator DNA, inducing a 120° global DNA bend towards the protein (Figure 17B) (Pérez-Martín, del Solar *et al.* 1989; Gomis-Rüth, Solá *et al.* 1998a). As observed in the co-crystals, two dimers would bind to the already described LSE and RSE subsites, and the other two to the LA and RA putative binding subsites, located laterally to the SE (del Solar, Pérez-Martín *et al.* 1990; Gomis-Rüth, Solá *et al.* 1998a). The binding of the two outer dimers seems to depend, directly or indirectly, on the nucleotide sequence, and also on the protein-protein interactions established between the dimers bound. In this extended complex, the repressor would cover not only the P_{cr} promoter -35 region, which is an RNA polymerase recognition sequence, but also the extended (*ext.*) -10 region and potential upstream promoter

elements (UP elements) (Figure 17A) (Estrem, Ross *et al.* 1999; Ross, Ernst *et al.* 2001). As a result, CopG would exert an enhanced repression of the P_{cr} promoter, accordingly with the observed tight control of pMV158 plasmid replication. The change in bending upon CopG binding would also explain that, besides hindering the access of the RNA polymerase to the promoter by steric exclusion, CopG is able to weaken the stable RNAP-P_{cr} complex and to dislodge the RNA polymerase from the promoter (Hernández-Arriaga, Rubio-Lepe *et al.* 2009).

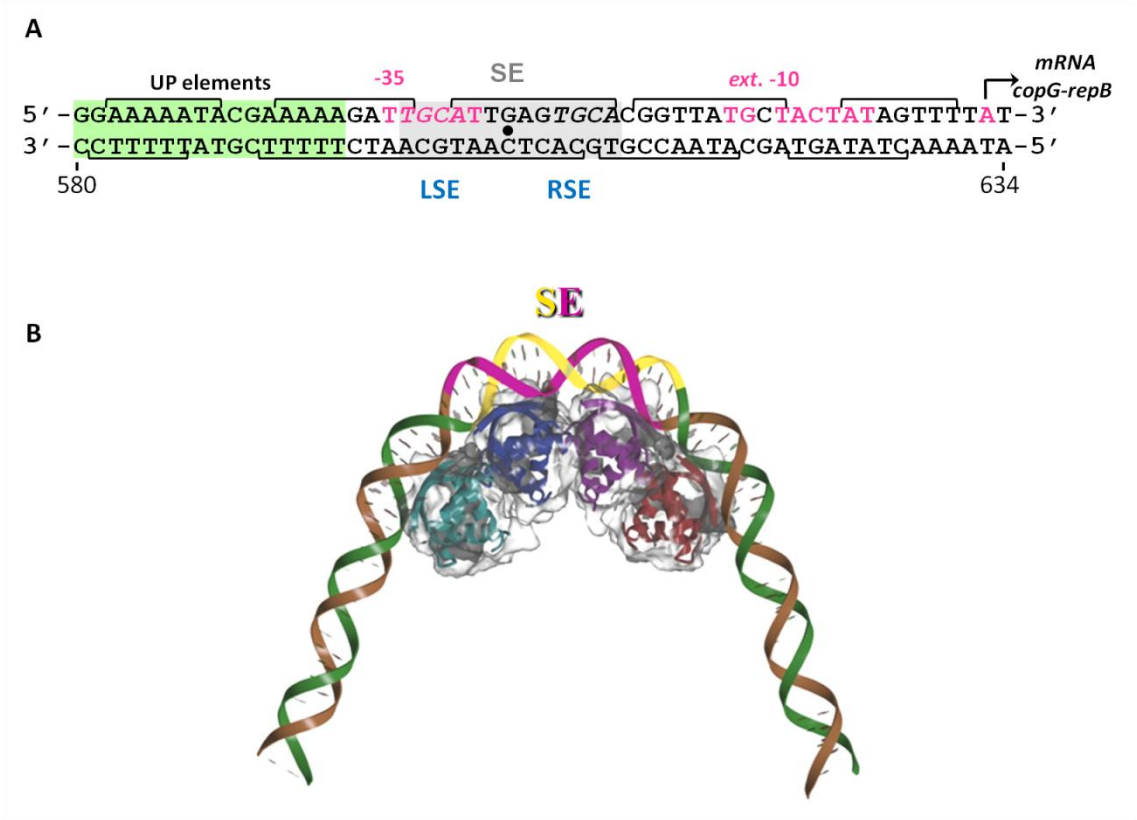


Figure 17. The DNA operator of CopG and the nucleoprotein complex model. (A) Sequence of the operator DNA of CopG corresponding to coordinates 580 to 634 of the pMV158 sequence. The 13-bp SE is *gray-shaded*, with its two-fold symmetry axis denoted by a *small circle*. The repeats within the SE are in *italics*. Promoter boxes -35 and *ext.* -10 and the start site of *copG-repB* operon transcription are in *pink*. The potential UP elements (proximal -46 to -38 and distal -57 to -47 (Estrem, Ross *et al.* 1999) are *green-shaded*. *Brackets* indicate the regions protected by CopG from DNase I cleavage. (B) Nucleoprotein complex model of CopG bound to its DNA operator. The DNA strands are displayed in *green* and *brown*, except the region corresponding to the SE, which is *yellow* and *magenta colored*. The structure of the CopG dimers is indicated by ribbon plots within a transparent Connolly surface. Four CopG dimers bind to four successive DNA helix turns. As a result, the target DNA is bent about 120°. Taken from (del Solar, Hernández-Arriaga *et al.* 2002).

Objectives

Within the RCR plasmids, which are commonly small, pMV158 is a good example to illustrate that economizing in genome-size is an important feature for this kind of replicons. CopG is the smallest transcriptional repressor described so far, conformed exclusively by the lowest structural unit that defines a RHH motif. Despite its smallness, CopG is able to repress the transcription from P_{cr} promoter with high affinity and high sequence specificity, owing to its highly cooperative binding properties. CopG is also the prototype of a growing family of repressor proteins involved in RC replication control, which contain the RHH motif. The study of the properties that underlie the high affinity and sequence specificity, and hence the efficient control exerted by CopG on the transcription, could be an important contribution to the basic knowledge of the protein-protein interactions that lie behind the cooperativity. Besides, it could establish the bases for designing useful gene-expression tools mainly for bacterial systems.

On this basis, and continuing the previous lab-work, our **HYPOTHESIS** consists of the following points:

- The functional nucleoprotein complex is formed by four CopG dimers bound to the promoter P_{cr}.
- The binding of CopG to its DNA operator is sequential and ordered.
- The binding specificity of CopG relies on the recognition of the nucleotide sequence of the various binding subsites as well as on the potential protein-protein interactions (cooperativity) between protein dimers bound to them.

For this thesis, it was proposed as **GENERAL OBJECTIVE**:
The study of the affinity and cooperative properties of the CopG repressor in the binding to its operator DNA.

Through the accomplishment of the following **PARTIAL OBJECTIVES**:

1. Analysis of the role of each of the operator subsites in binding affinity of CopG.
2. Study of binding affinity and cooperativity of three protein variants (wild-type and two mutants) for distinct operators carrying modifications both in the nucleotide sequence of the subsites and in subsite spacing.
3. Structural resolution of CopG (wild-type and mutants) in solution, both free and bound to DNA by NMR. This has been performed in collaboration with Oscar Millet's group from the Center for Cooperative Research in Biosciences (CIC bioGUNE) in Bilbao, Vizcaya.

Materials and Methods

This chapter is arranged into two main blocks: i) *chemical and biological material acquisition* and ii) *experimental procedures*. In the former, all the chemical or living substrates that are necessary to execute the experimentation are detailed, as well as the basic techniques and processes pointed to produce them. Trade-marks and product specifications are included in the text. The latter is an extended section about the experimental approaches that were designed and performed for fulfilling the objectives.

Part I. CHEMICAL AND BIOLOGICAL MATERIAL ACQUISITION

1. Bacterial strains and manipulation

Genotypic characteristics and origin of the bacterial strains employed in this work are listed in Table 1. *E. coli* M15 [pREP4] was used to overproduce CopG proteins, and *S. pneumoniae* 708 to obtain pLS1, pLS1*copG7* and pLS1*copG8* plasmid DNA as well as the pC194-based recombinant plasmids. The pneumococcal strain was also used to perform the *in vivo* incompatibility experiments.

Table 1 | Bacterial strains

Strains	Genotype	Source/Reference
E. coli M15 [pREP4]	Strain derived from *E.coli* K12, carrying the deletion M15 in the *lacZ* gene	Purchased from QIAGEN (Villarejo & Zabin 1974)
S. pneumoniae 708	*end-1, exo-2, trt-1, hex-4, malM594*	(Lacks & Greenberg 1977)

1.1. Culture media

E. coli M15 [pREP4] cells containing the recombinant plasmids derived from the pQE-1 vector (part I, 3.1) were grown in LB medium (Luria & Burrous 1957) with selection for ampicillin (Amp, 100 μg/ml) and kanamycin (Km, 25 μg/ml) resistance. Liquid cultures were incubated with vigorous shaking under airy conditions, and the bacterial growth was controlled by measuring the absorbance at 600 nm. In the case of LB solid media (supplemented with 1.5% agar), cells were grown at the surface.

S. pneumoniae cells were grown in AGCH medium (Lacks 1966) supplemented with 0.3% sucrose and 0.2% yeast extract, with selection for resistance to tetracycline (Tc, 1 μg/ml) for pMV158-based plasmids or to chloramphenicol (Cm, 3 μg/ml) for pC194-based replicons. Optimum growth of pneumococcal cells is obtained under a low aeration conditions, so liquid cultures

were kept static, and for the solid media, an upper layer of medium was added covering the basal one where the cells were inoculated. Liquid cultures were monitored by measurements of absorbance at 650 nm. Incubation temperature for both *E. coli* and *S. pneumoniae* was 37 °C. A spectrophotometer Spectronic 20D was utilized for control of turbidity. All strains were stored at -80 °C with 10% glycerol.

1.2. Transformation

E. coli electro-transformation was carried out as described (Dower, Miller *et al.* 1988). Competent cells were transferred to an electroporation cuvette and given an electric pulse under published conditions (25 µF, 2.5 kV/cm and 200 Ω) in a Gene pulser™ (Bio-Rad). Immediately, bacteria were inoculated in SOC medium (2% Bactotryptone 0.5% yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose). Cells were incubated for 1 h at 37 °C for recuperation, and then plated on the appropriate selection medium to screen for transformants.

Transformation of *S. pneumoniae* was performed basically as described (Balganesh & Lacks 1984). Competent pneumococcal cells were inoculated in prewarmed AGCH medium (Lacks 1966), supplemented with 0.1% sucrose plus 0.1% $CaCl_2$, and incubated for 20 min at 30 °C. Donor DNA (pC194-based recombinant plasmids) was mixed with 1 ml of culture and kept for 40 min at 30 °C for natural transformation. Then, the mixtures were incubated at 37 °C for 70 min before inducing expression of the *cat* gene (which encodes the chloramphenicol acetyl transferase) by addition of subinhibitory concentrations of the antibiotic (Cm, 0.5 µg/ml) and incubation for 20 more min at 37 °C. Transformants were selected in solid medium containing 3 µg/ml Cm.

2. DNA preparations

2.1. Plasmids

Details of the plasmids utilized in this thesis are in Table 2.

Table 2 | Plasmids

Plasmids	Size (bp)	Description[a]	Source/Reference
pMV158	5540	Isolated from a clinical isolate of *S. agalactiae* MV158; Tc^R	(Burdett 1980)
pLS1	4408	pMV158 derivative (Δ*mobM,* Δ*ssoU*); Tc^R	(Lacks, López *et al.* 1986)

pLS1*copG7*	4408	pLS1 mutant, having a C to A (743) transversion in *copG* (*copG7*); Tc^R	(del Solar, de la Campa *et al.* 1989)
pLS1*copG8*	4408	pLS1 mutant, having a G to A (728) transition in *copG* (*copG8*); Tc^R	(Acebo, Alda *et al.* 1996)
pQE-1	3466	Expression vector; Amp^R	QIAGEN
pQE-1*copG*	3604	Recombinant with *copG* cloned in the single *Puv*II site of pQE-1; Amp^R	This study
pQE-1*copG7*	3604	Recombinant with *copG7* cloned in the single *Puv*II site of pQE-1; Amp^R	This study
pQE-1*copG8*	3604	Recombinant with *copG8* cloned in the single *Puv*II site of pQE-1; Amp^R	This study
pC194	2910	Isolated from *S. aureus*; Cm^R	(Horinouchi & Weisblum 1982)
pC194-sLSE	2965	Recombinant carrying a copy of super-operator sLSE cloned in the single *Hin*dIII site of pC194; Cm^R	This study
pC194-sLSE(2)	3024	Recombinant carrying two copies of super-operator sLSE cloned in the single *Hin*dIII site of pC194; Cm^R	This study
pC194-sRA3	2965	Recombinant carrying a copy of super-operator sRA3 cloned in the single *Hin*dIII site of pC194; Cm^R	This study

[a] *Tc^R , Amp^R, Cm^R,* resistance to tetracycline, ampicillin and chloramphenicol, respectively

2.1.1 Extraction of plasmids

Plasmid DNA from *E. coli* was extracted and purified by the FavorPrep™ Plasmid DNA MiniExtraction Kit (FAVORGEN). This commercial kit is based on the alkaline lysis method described by (Bimboim & Doly 1979). A modification of this method (Stassi, Lopez *et al.* 1981) was followed to prepare purified plasmids from *S. pneumoniae*. Pneumococcal crude extracts were prepared as published (del Solar, Puyet *et al.* 1987).

2.2. DNA fragments

Two DNA fragments were employed throughout the experimental work. Table 3.

Table 3 | DNA fragments

DNA fragment	Description	Coordinates of pMV158 sequence
138-bp	*copG* gene	655-792
239-bp	CopG operator	489-727

2.2.1. Obtainment of the 138-bp DNA fragment containing the *copG* gene

Genes encoding $CopG_{WT}$, $CopG_{G25E}$ and $CopG_{A30E}$ proteins (along with its natural TAA stop codon in the pMV158 sequence) were amplified by PCR using Phusion® DNA Polymerase (Finnzymes), the synthetic oligonucleotides *Forward* 655-677 and *Reverse* 792-770 (Table 4) as primers, and pLS1, pLS1*copG8* and pLS1*copG7* DNAs, respectively, as templates. The primers were previously phosphorylated using the T4-Polynucleotide kinase (T4-PNK) (New England BioLabs) following the manufacturer's instructions. Amplification was performed in 50-100 µl of the Phusion® HF Buffer provided with the enzyme. Each component was added at the recommended concentration. A preheating step of 30 s at 98 °C was performed, followed by 25 amplification cycles of hybridization/elongation whose temperature conditions were 20 s at 55 °C and 5 s at 72 °C.

2.2.2. Obtainment of the 239-bp DNA fragment containing the CopG operator

A DNA fragment containing the wild-type CopG operator was amplified by PCR using Phusion® DNA polymerase (Finnzymes) and pLS1 DNA as template. Synthetic oligonucleotides *Forward* 489-505 and *Reverse* 727-712 (Table 4) were employed as primers. The reaction was carried out in a final volume of 50-100 µl of Phusion® HF Buffer supplemented with all other required components at the recommended concentration. Twenty-five amplification cycles of hybridization/elongation were carried out (30 s at 52.5 °C and 30 s at 72 °C), after an initial denaturation step of 30 s at 98 °C.

Table 4 | Primers

Name	Size (nt)	Sequence 5'- 3'	IDT Data		Application
			nmol/ OD_{260}	µg/ OD_{260}	
Forward 489-505[a]	17	CGCCTTTAGCCTTAGAG	6.39	33	PCR
Reverse 727-712[a]	16	CCATCTCTCTTGCCAT	7.29	34.57	
Forward 655-677[a]	23	ATGAAAAAAAGATTGACGATAAC	3.97	28.31	PCR/ Sequencing
Reverse 792-770[a]	23	TTATTTTTCTTGACCTTTCTTGT	5.05	35.08	
Forward 3375-3392[b]	18	CCCGAAAAGTGCCACCTG	5.93	32.33	PCR/ Sequencing
Reverse 248-230[b]	19	GTTCTGAGGTCATTACTGG	5.52	32.3	
Reverse 2534-2518[c]	17	ACTGGGCAGTGTCTTAA	6.01	31.4	Sequencing

These primers hybridize with plasmids pMV158[a], pQE-1[b] and pC194[c]. *IDT*, Integrated DNA Technologies (Oligo-analyzer facility, on-line)

2.2.3. Purification of DNA fragments

DNA fragments obtained by PCR were purified, according to their size, from 8% or 5% polyacrylamide (PAA) native gels (138- and 239-bp fragments respectively). PCR mixtures were directly mixed with 1X BXGE loading buffer (0.025% bromophenol blue, 0.025% xylene cyanol, 6% glycerol, 1 mM EDTA) and loaded onto preparative gels. Electrophoresis was run at 8 V/cm in 1X TBE buffer (89 mM Tris base pH 8.3, 89 mM boric acid, EDTA 2.5 mM).

The bands corresponding to the amplification products were stained with ethidium bromide (EtBr), visualized under UV-radiation and excised with a scalpel-blade. DNA fragments were recovered through incubation in elution buffer (20 mM Tris-HCl pH 8.0, 200 mM NaCl, 1 mM EDTA) overnight at 42 °C with shaking. A 1-h second elution was carried out to increase the DNA recovery. To eliminate the PAA rubbish, samples were filtrated on "Centrifuge Tube Filter" (Spin X, COSTAR) and the aqueous solution was concentrated by extraction with isobutanol. Co-precipitant Pellet Paint™ was added as a carrier for ethanol precipitation. Finally, pellets were dissolved in Milli-Q water and the DNA concentration was measured on agarosa gels using a control of known concentration, or by 260 nm absorbance in a NanoDrop™ Spectrophotometer when samples lacked Pellet Paint™. DNA fragments were stored at -80 °C.

The procedure to obtain the radiolabeled 239-bp fragment is described in 2.4.2 of this part.

2.3. Oligonucleotides

Chemical synthesis of a pool of ssDNAs varying in sequence and length was required for most of the experiments in this work. All of them were synthesized and purified by HPLC in the Protein Chemistry Facility at CIB. Oligonucleotides were subsequently separated and purified from denaturing PAA sequencing gels in our lab, as described below. These oligonucleotides, with their names and sequences as well as the data used to calculate their concentration, are listed in Table 5. All of them were annealed with their complementary strands (A + B) as described in 2.3.2, in order to obtain distinct operator variants.

Table 5 | Oligonucleotides

Name	Size (nt)	Sequence 5'- 3'	IDT Data	
			nmol/OD_{260}	µg/OD_{260}
o(RSE) A	18	ATTGAGTGCACGGTTATG	5.6	31.17
o(RSE) B	18	CATAACCGTGCACTCAAT	5.82	31.57
o(RSE*) A	18	ATTGAGTGCACGCTGCAG	5.77	31.97
o(RSE*) B	18	CTGCAGCGTGCACTCAAT	6.07	33.15
o(LSE-RSE) A	21	AGATTGCATTGAGTGCACGGT	4.83	31.38
o(LSE-RSE) B	21	ACCGTGCACTCAATGCAATCT	5.08	32.25
o(sLSE-RSE) A	21	AGAGTGCATTGAGTGCACGGT	4.77	31.11
o(sLSE-RSE) B	21	ACCGTGCACTCAATGCACTCT	5.23	33.1
o(sLSE+1-RSE) A	22	AGAGTGCATTTGAGTGCACGGT	4.59	31.35
o(sLSE+1-RSE) B	22	ACCGTGCACTCAAATGCACTCT	4.92	32.69
o(RSE-RA) A	25	TTGAGTGCACGGTTATGCTACTATA	4.11	31.59
o(RSE-RA) B	25	TATAGTAGCATAACCGTGCACTCAA	3.98	30.4
o(RSE-sRA1) A	25	TTGAGTGCACGGTTATGCACCTATA	4.13	31.69
o(RSE-sRA1) B	25	TATAGGTGCATAACCGTGCACTCAA	4.05	30.95
o(RSE-sRA1+1) A	25	TTGAGTGCACGGTTATTGCACCTAT	4.25	32.54
o(RSE-sRA1+1) B	25	ATAGGTGCAATAACCGTGCACTCAA	3.98	30.49
o(LA-LSE) A	25	AAAATACGAAAAAGATTGCATTGAG	3.71	28.75
o(LA-LSE) B	25	CTCAATGCAATCTTTTTCGTATTTT	4.42	33.43
o(sLA1-LSE) A	25	AAAATGCACAAAAGATTGCATTGAG	3.81	29.43
o(sLA1-LSE) B	25	CTCAATGCAATCTTTTGTGCATTTT	4.43	33.6

Name	Size (nt)	Sequence 5'- 3'	IDT Data	
			nmol/OD_{260}	µg/OD_{260}
o(sLA1-1-LSE) A	25	AAAAATGCACAAAGATTGCATTGAG	3.81	29.43
o(sLA1-1-LSE) B	25	CTCAATGCAATCTTTGTGCATTTTT	4.43	33.6
o(RSE-sRA3) A	25	TTGAGTGCACGGTTATGTGCATATA	4.04	31.22
o(RSE-sRA3) B	25	TATATGCACATAACCGTGCACTCAA	4.07	30.87
o(RSE-sRA3-1) A	25	TTGAGTGCACGGTTAGTGCATATAG	4.02	31.13
o(RSE-sRA3-1) B	25	CTATATGCACTAACCGTGCACTCAA	4.18	31.66
WT A	55	GGAAAAATACGAAAAAGATTGCATTGAGTGCACGGTTATGCTACTATAGTTTTAT	1.79	30.5
WT B	55	ATAAAACTATAGTAGCATAACCGTGCACTCAATGCAATCTTTTTCGTATTTTTCC	1.89	31.68
RSE$^-$ A	55	GGAAAAATACGAAAAAGATTGCATTGGACTTGAGGTTATGCTACTATAGTTTTAT	1.79	30.49
RSE$^-$ B	55	ATAAAACTATAGTAGCATAACCTCAAGTCCAATGCAATCTTTTTCGTATTTTTCC	1.87	31.45
LSE$^-$ A	55	GGAAAAATACGAAAAAGATCAAGTCGAGTGCACGGTTATGCTACTATAGTTTTAT	1.77	30.2
LSE$^-$ B	55	ATAAAACTATAGTAGCATAACCGTGCACTCGACTTGATCTTTTTCGTATTTTTCC	1.91	32.02
RA$^-$ A	55	GGAAAAATACGAAAAAGATTGCATTGAGTGCACGCTGCAGCCCGGGGGATCCACT	1.83	31.28
RA$^-$ B	55	AGTGGATCCCCCGGGCTGCAGCGTGCACTCAATGCAATCTTTTTCGTATTTTTCC	2.01	33.71
LA$^-$ A	55	TTGATATCGAATTCCTAATTGCATTGAGTGCACGGTTATGCTACTATAGTTTTAT	1.88	31.86
LA$^-$ B	55	ATAAAACTATAGTAGCATAACCGTGCACTCAATGCAATTAGGAATTCGATATCAA	1.77	29.97
LSE-RSE A	55	TTGATATCGAATTCCTAATTGCATTGAGTGCACGCTGCAGCCCGGGGGATCCACT	1.93	32.72
LSE-RSE B	55	AGTGGATCCCCCGGGCTGCAGCGTGCACTCAATGCAATTAGGAATTCGATATCAA	1.87	31.77
LSE-RSE* A	55	TTGATATCGAATTCCTAATTGCATTGAGTGCACGGTTATGCCCGGGGGATCCACT	1.91	32.43
LSE-RSE* B	55	AGTGGATCCCCCGGGCATAACCGTGCACTCAATGCAATTAGGAATTCGATATCAA	1.85	31.28
RSE-RA A	55	TTGATATCGAATTCCTAATCAAGTCGAGTGCACGGTTATGCTACTATAGTTTTAT	1.86	31.52
RSE-RA B	55	ATAAAACTATAGTAGCATAACCGTGCACTCGACTTGATTAGGAATTCGATATCAA	1.79	30.27

Name	Size (nt)	Sequence 5'- 3'	IDT Data	
			nmol/OD_{260}	µg/OD_{260}
LA-LSE A	55	GGAAAAATACGAAAAAGATTGCATTGGACTTGAGCTGCAGCCCGGGGGATCCACT	1.83	31.27
LA-LSE B	55	AGTGGATCCCCCGGGCTGCAGCTCAAGTCCAATGCAATCTTTTTCGTATTTTTCC	1.99	33.45
LSE$^+$ A	55	TTGATATCGAATTCCTAATTGCATTGGACTTGAGCTGCAGCCCGGGGGATCCACT	1.93	32.7
LSE$^+$ B	55	AGTGGATCCCCCGGGCTGCAGCTCAAGTCCAATGCAATTAGGAATTCGATATCAA	1.86	31.53
RSE$^+$ A	55	TTGATATCGAATTCCTAATCAAGTCGAGTGCACGCTGCAGCCCGGGGGATCCACT	1.91	32.37
RSE$^+$ B	55	AGTGGATCCCCCGGGCTGCAGCGTGCACTCGACTTGATTAGGAATTCGATATCAA	1.89	32.1
LA$^+$ A	55	GGAAAAATACGAAAAAGATCAAGTCGGACTTGAGCTGCAGCCCGGGGGATCCACT	1.81	30.95
LA$^+$ B	55	AGTGGATCCCCCGGGCTGCAGCTCAAGTCCGACTTGATCTTTTTCGTATTTTTCC	2.01	33.82
RA$^+$ A	55	TTGATATCGAATTCCTAATCAAGTCGGACTTGAGGTTATGCTACTATAGTTTTAT	1.86	31.51
RA$^+$ B	55	ATAAAACTATAGTAGCATAACCTCAAGTCCGACTTGATTAGGAATTCGATATCAA	1.78	30.05
NS A	55	TTGATATCGAATTCCTAATCAAGTCGGACTTGAGCTGCAGCCCGGGGGATCCACT	1.91	32.35
NS B	55	AGTGGATCCCCCGGGCTGCAGCTCAAGTCCGACTTGATTAGGAATTCGATATCAA	1.88	31.86
sLSE A	55	GGAAAAATACGAAAAAGAGTGCATTGAGTGCACGGTTATGCTACTATAGTTTTAT	1.78	30.4
sLSE B	55	ATAAAACTATAGTAGCATAACCGTGCACTCAATGCACTCTTTTTCGTATTTTTCC	1.91	31.99
sRA1 A	55	GGAAAAATACGAAAAAGATTGCATTGAGTGCACGGTTATGCACCTATAGTTTTAT	1.79	30.56
sRA1 B	55	ATAAAACTATAGGTGCATAACCGTGCACTCAATGCAATCTTTTTCGTATTTTTCC	1.9	31.95
sRA2 A	55	GGAAAAATACGAAAAAGATTGCATTGAGTGCACGGTTATGCCACTATAGTTTTAT	1.79	30.56
sRA2 B	55	ATAAAACTATAGTGGCATAACCGTGCACTCAATGCAATCTTTTTCGTATTTTTCC	1.9	31.95
sRA3 A	55	GGAAAAATACGAAAAAGATTGCATTGAGTGCACGGTTATGTGCATATAGTTTTAT	1.78	30.36
sRA3 B	55	ATAAAACTATATGCACATAACCGTGCACTCAATGCAATCTTTTTCGTATTTTTCC	1.91	31.91

Name	Size (nt)	Sequence 5'- 3'	IDT Data	
			nmol/OD_{260}	µg/OD_{260}
sRA4 A	55	GGAAAAATACGAAAAAGATTGCATTGAGTGCACGGTTATGTGACTATAGTTTTAT	1.78	30.36
sRA4 B	55	ATAAAACTATAGTCACATAACCGTGCACTCAATGCAATCTTTTTCGTATTTTTCC	1.9	31.81
sLA1 A	55	GGAAAAATGCACAAAAGATTGCATTGAGTGCACGGTTATGCTACTATAGTTTTAT	1.81	30.84
sLA1 B	55	ATAAAACTATAGTAGCATAACCGTGCACTCAATGCAATCTTTTGTGCATTTTTCC	1.89	31.75
sLA2 A	55	GGAAAAAGTGCAAAAAGATTGCATTGAGTGCACGGTTATGCTACTATAGTTTTAT	1.8	30.74
sLA2 B	55	ATAAAACTATAGTAGCATAACCGTGCACTCAATGCAATCTTTTTGCACTTTTTCC	1.9	31.95

2.3.1. Purification of oligonucleotides

Sequencing gels containing 8 M urea and PAA concentrations fluctuating in a range from 8% to 16% (depending on the oligonucleotide size) were used. Table 6 shows the components employed to prepare the denaturing gels used both in purifying oligonucleotides and in resolving footprinting reactions (part II, 5). Gel solutions were poured into the assembled Sequi-Gen®GT system (Bio-Rad).

Table 6 |Components of denaturing gels

Reagents	**8%**	**12%**	**15%**	**16%**
40% Acryl:Bis (19:1) (ml)	*10*	*15*	*18.75*	*20*
10X TBE (ml)	*5*	*5*	*5*	*5*
Urea (g)	*24.02*	*24.02*	*24.02*	*24.02*
Polymerize by adding:				
APS (10%) (µl)	*300*	*300*	*300*	*300*
TEMED (µl)	*25*	*25*	*25*	*25*
Milli-Q water added to	*50 ml*			

Gels were pre-run for at least 1 h at a constant power of 45-55 W, until the gel temperature reached 45-50 °C. Meanwhile, samples were vigorously mixed with 1 volume of deionized formamide, heated for 5 min at 55 °C and cooled on ice until they were loaded onto the sequencing gel. Approximately two OD_{260} units of ssDNA were loaded per well. The same volume of a dye-containing loading buffer was loaded in a separate lane to act as a migration tracker. Electrophoresis was continued under the same conditions. To visualize the DNA bands, the gel was covered with plastic Saran™ wrap, then carefully peeled off from the glass plate and placed onto Fluor-Coated TLC plates (Ambion). Oligonucleotide bands were detected under short-wavelength UV radiation, excised with a scalpel-blade and immediately put into Eppendorf tubes.

The ssDNA was eluted from the PAA slices overnight at 37 °C with shaking in elution buffer (10 mM magnesium acetate, 200 mM NaCl, 0.1% SDS). A 2-h second elution was carried out looking for increased DNA recovery. PAA rubbish was removed by filtration in “Micro-Centrifuge Tube Filter” (Spin X, COSTAR). Thereafter, samples were concentrated with 1-butanol, and ethanol precipitated. ssDNA pellets were dissolved in Milli-Q water and purified by liquid chromatography using the columns Micro-Spin™ G25 (Amersham) according to manufacturer’s instructions. Final concentration was determined from absorbance at 260 nm (measured in a NanoDrop™ Spectrophotometer), using the sequence-dependent data obtained from the on-

line Oligo-Analyzer facility from Integrated DNA Technologies (IDT); these data are included in Table 5. All the oligonucleotides recovered by this procedure were annealed, regardless they were previously subjected to ^{32}P-labeling or not.

2.3.1.1. Purification of radiolabeled oligonucleotides

Albeit this procedure was not a routinely practice in this thesis, sometimes gel-purification of ^{32}P-labeled oligonucleotides was required. This was the case with the 55-bp oligonucleotides WTA and WTB that were to be used in footprinting experiments. This second purification, performed after labeling of the oligonucleotide, pursued total removal of smaller-than-full-length molecules, which could impair footprinting resolution.

The procedure was as follows. Once the labeled oligonucleotides were precipitated and dissolved in Milli-Q water, 1 volume of deionized formamide was added and thoroughly mixed. Samples were heated for 5 min at 55 °C and hold on ice until being loaded on a denaturing gel (8% PAA, 8 M urea). Electrophoresis was run as indicated above for oligonucleotide purification. The gel was then transferred to a discarded X-ray film, covered with plastic Saran™ wrap, and autoradiographed by exposition to X-ray film (Curix RP2 Plus, AGFA) no longer than 10 s. After outlining the bands in the developed film, this was superposed on the gel and aligned with the fluorescents rulers located beside it. Section corresponding to the labeled oligonucleotide was excised using a scalpel-blade and ssDNA was recovered from the PAA slices by elution in buffer (500 mM ammonium acetate, 10 mM magnesium acetate, 0.1% SDS), overnight at 37 °C with shaking. A 2-h second elution was also performed to increase recovery. In order to eliminate residual PAA fragments, samples were filtrated through the “Micro-Centrifuge Tube Filter” (Spin X, COSTAR). Co-Precipitant Pellet-Paint™ and 0.1 volume of 3 M sodium acetate were added before ethanol precipitation. Labeled-ssDNA was finally dissolved in Milli-Q water and led to the annealing reaction.

2.3.2. ssDNA annealing

Pairs of non-radioactive complementary oligonucleotides were annealed at equimolar concentrations in hybridization buffer (2 mM Tris-HCl pH 8.0, 100 mM NaCl, 0.2 mM EDTA). Samples were denatured for 10 min at 90 °C and slowly cooled to room temperature. Reaction volumes ranged from 25 µl to 30 µl, and final concentrations of double stranded DNA (dsDNA) were from 3.6 µM to 12 µM. When preparing radiolabeled dsDNA, the labeled strand was incubated with a 3-fold molar excess of the unlabeled complementary strand in a reaction volume of 25 µl of the same hybridization buffer. Final concentrations of labeled-dsDNA ranged

from 0.4 µM to 2 µM. Non-radioactive samples were stored at -80 °C and those labeled were kept at -20 °C. All the dsDNA employed in this thesis are described in Table 7. See Figure 18 for definition of the different elements of the CopG operator DNA.

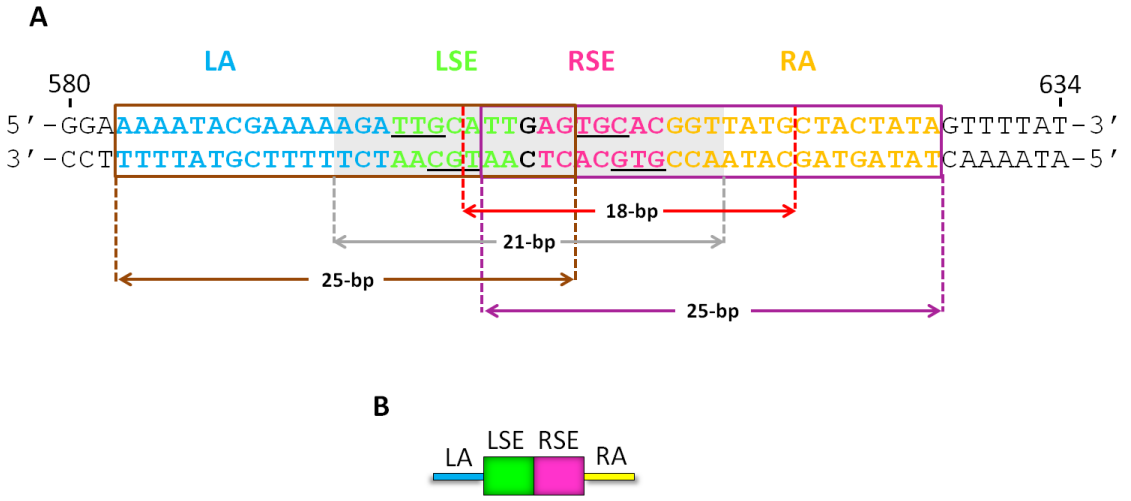


Figure 18. Arbitrary definition of the protein binding subsites in the operator of CopG. (A) The two crystallographically-defined central subsites as well as the two putative lateral binding subsites are indicated on the WT 55-bp oligonucleotide spanning coordinates 580 to 634 of the pMV158 sequence. The LSE and RSE are depicted in *green* and *pink letters* respectively. Arbitrarily, these subsites were defined to consist of the 7-bp sequences located on both sides of the G:C central base-pair (*black in bold*). This G:C base-pair represents the symmetric axis of the SE. Bases contacted by CopG in the 19-bp dsDNA co-crystals are *underlined* in the LSE and RSE. *Blue* and y*ellow letters* show the regions likely to contain the recognition sequences of the protein in the putative binding sites named LA and RA, respectively. *Violet box* indicates the 25-bp oligonucleotide corresponding to the operator right half, which encompasses the RSE and the putative RA binding site. *Brown box* indicates the 25-bp oligonucleotide corresponding to the operator left half, which comprises the LSE and the putative LA binding site. *Gray-shaded* sequence indicates the 21-bp oligonucleotide corresponding to the operator central region, which includes the LSE and RSE subsites plus three adjacent bases on each side. *Red dashed lines* indicate the 18-bp oligonucleotide spanning the isolated RSE plus the adjacent bases. (B) Scheme employed along the thesis for representing the subsites in the operator.

Table 7 | Description of the dsDNAs used in this thesis[a].

Name	Size	Sequence	Description
o(RSE)	18	5′-ATT**G**AGTGCACGGTTATG-3′ 3′-TAA**C**TCACGTGCCAATAC-5′	Oligonucleotide (602-619) comprising exclusively the RSE and adjacent bases.
o(RSE*)	18	5′-ATT**G**AGTGCACG**C**T**GCA**G-3′ 3′-TAA**C**TCACGTGC**G**A**CGT**C-5′	Oligonucleotide (602-619) comprising exclusively the RSE plus adjacent bases, but changing G(614) and from T(616) to T(618) by the sequence in LSE-RSE operator.
o(LSE-RSE)	21	5′-AGATTGCATT**G**AGTGCACGGT-3′ 3′-TCTAACGTAA**C**TCACGTGCCA-5′	Oligonucleotide (595-615) consisting exclusively of the LSE and RSE plus adjacent bases.
o(sLSE-RSE)	21	5′-AGA**G**TGCATT**G**AGTGCACGGT-3′ 3′-TCT**C**ACGTAA**C**TCACGTGCCA-5′	Oligonucleotide (595-615) consisting exclusively of the LSE modified as in the sLSE oligonucleotide, and the RSE plus adjacent bases. See description of sLSE.
o(sLSE+1-RSE)	22	5′-AGA**G**TGCAT**T**T**G**AGTGCACGGT-3′ 3′-TCT**C**ACGTA**A**A**C**TCACGTGCCA-5′	Oligonucleotide comprising exclusively the LSE modified as in the sLSE oligonucleotide, and the RSE plus adjacent bases, with an extra T between positions 603 and 604 of the operator sequence. See description of sLSE.
o(RSE-RA)	25	5′-TT**G**AGTGCACGGTTATGCTACTATA-3′ 3′-AA**C**TCACGTGCCAATACGATGATAT-5′	Oligonucleotide (603-627) consisting exclusively of the RSE plus adjacent bases and the putative RA binding site.
o(RSE-sRA1)	25	5′-TT**G**AGTGCACGGTTATGC**AC**CTATA-3′ 3′-AA**C**TCACGTGCCAATACG**TG**GATAT-5′	Oligonucleotide (603-627) consisting exclusively of the RSE plus adjacent bases and the putative RA binding site modified as in the sRA1 oligonucleotide. See description of sRA1.

Name	Size	Sequence	Description
o(RSE+1-sRA1)	25	5′-TT**G**AGTGCACGGTTA**T**TGC**AC**CTAT-3′ 3′-AA**C**TCACGTGCCAAT**A**ACG**TG**GATA-5′	Oligonucleotide comprising exclusively the RSE plus adjacent bases and the putative RA binding site modified as in the sRA1 oligonucleotide, with an extra T between position 617 and 618 of the operator sequence. See description of sRA1.
o(RSE-sRA3)	25	5′-TT**G**AGTGCACGGTTATG**TGCA**TATA-3′ 3′-AA**C**TCACGTGCCAATAC**ACGT**ATAT-5′	Oligonucleotide (603-627) consisting exclusively of the RSE plus adjacent bases and the putative RA binding site modified as in the sRA3 oligonucleotide. See description of sRA3.
o(RSE-1-sRA3)	25	5′-TT**G**AGTGCACGGTTA|G**TGCA**TATAG-3′ 3′-AA**C**TCACGTGCCAAT|C**ACGT**ATATC-5′	Oligonucleotide consisting exclusively of the RSE plus adjacent bases and the putative RA binding site modified as in the sRA3 oligonucleotide with a deletion of T(618) of the operator sequence (*vertical line*). See description of sRA3.
o(sLA1-LSE)	25	5′-AAAAT**G**C**AC**AAAAGATTGCATT**G**AG-3′ 3′-TTTTA**C**G**TG**TTTTCTAACGTAA**C**TC-5′	Oligonucleotide (583-607) consisting only of the putative LA binding site modified as in the sLA1 oligonucleotide, and the LSE plus adjacent bases. See description of sLA1.
o(sLA1-1-LSE)	25	5′-AAAAAT**G**C**AC**AAA|GATTGCATT**G**AG-3′ 3′-TTTTTA**C**G**TG**TTT|CTAACGTAA**C**TC-5′	Oligonucleotide consisting only of the putative LA binding site modified as in the sLA1 oligonucleotide, and the LSE plus adjacent bases, with a deletion of A(595) of the operator sequence (*vertical line*).
o(LA-LSE)	25	5′-AAAATACGAAAAAGATTGCATT**G**AG-3′ 3′-TTTTATGCTTTTTCTAACGTAA**C**TC-5′	Oligonucleotide (583-607) consisting exclusively of the putative LA binding site and the LSE plus adjacent bases.

Name	Size	Sequence	Description
WT	55	5′-GGAAAAATACGAAAAAGATTGCATT**G**AGTGCACGGTTATGCTACTATAGTTTTAT-3′ 3′-CCTTTTTATGCTTTTTCTAACGTAA**C**TCACGTGCCAATACGATGATATCAAAATA-5′	Oligonucleotide containing the full-length wild-type operator: putative LA, LSE, RSE and putative RA binding sites (580-634).
RSE⁻	55	5′-GGAAAAATACGAAAAAGATTGCATT**G**GACTTGAGGTTATGCTACTATAGTTTTAT-3′ 3′-CCTTTTTATGCTTTTTCTAACGTAA**C**CTGAACTCCAATACGATGATATCAAAATA-5′	Oligonucleotide containing the full-length operator but lacking the RSE (mutated sequence: 606-612).
LSE⁻	55	5′-GGAAAAATACGAAAAAGATCAAGTC**G**AGTGCACGGTTATGCTACTATAGTTTTAT-3′ 3′-CCTTTTTATGCTTTTTCTAGTTCAG**C**TCACGTGCCAATACGATGATATCAAAATA-5′	Oligonucleotide containing the full-length operator but lacking the LSE (mutated sequence: 598-604).
RA⁻	55	5′-GGAAAAATACGAAAAAGATTGCATT**G**AGTGCACGCTGCAGCCCGGGGGATCCACT-3′ 3′-CCTTTTTATGCTTTTTCTAACGTAA**C**TCACGTGCGACGTCGGGCCCCCTAGGTGA-5′	Oligonucleotide containing the full-length operator but lacking the region containing the putative RA binding site (mutated sequence: 614-633).
LA⁻	55	5′-TTGATATCGAATTCCTAATTGCATT**G**AGTGCACGGTTATGCTACTATAGTTTTAT-3′ 3′-AACTATAGCTTAAGGATTAACGTAA**C**TCACGTGCCAATACGATGATATCAAAATA-5′	Oligonucleotide containing the full-length operator but lacking the region containing the putative LA binding site (mutated sequence: 580-596).
RSE⁺	55	5′-TTGATATCGAATTCCTAATCAAGTC**G**AGTGCACGCTGCAGCCCGGGGGATCCACT-3′ 3′-AACTATAGCTTAAGGATTAGTTCAG**C**TCACGTGCGACGTCGGGCCCCCTAGGTGA-5′	Oligonucleotide containing the full-length operator but conserving only the RSE plus the 1-bp adjacent on the right (wild-type sequence: 605-613).
LSE⁺	55	5′-TTGATATCGAATTCCTAATTGCATT**G**GACTTGAGCTGCAGCCCGGGGGATCCACT-3′ 3′-AACTATAGCTTAAGGATTAACGTAA**C**CTGAACTCGACGTCGGGCCCCCTAGGTGA-5′	Oligonucleotide containing the full-length operator but conserving only the LSE plus the 1-bp adjacent on the left (wild-type sequence: 597-605).
RA⁺	55	5′-TTGATATCGAATTCCTAATCAAGTCGGACTTGAGGTTATGCTACTATAGTTTTAT-3′ 3′-AACTATAGCTTAAGGATTAGTTCAGCCTGAACTCCAATACGATGATATCAAAATA-5′	Oligonucleotide containing the full-length operator but conserving exclusively the region containing the putative RA binding site (wild-type sequence: 613-634).
LA⁺	55	5′-GGAAAAATACGAAAAAGATCAAGTCGGACTTGAGCTGCAGCCCGGGGGATCCACT-3′ 3′-CCTTTTTATGCTTTTTCTAGTTCAGCCTGAACTCGACGTCGGGCCCCCTAGGTGA-5′	Oligonucleotide containing the full-length operator but conserving exclusively the region containing the putative LA binding site (wild-type sequence: 580-597).

Name	Size	Sequence	Description
LSE-RSE	55	5′-**TTGATATCGAATTCCTA**ATTGCATT**G**AGTGCACG**CTGCAGCCCGGGGGATCCAC**T-3′ 3′-**AACTATAGCTTAAGGAT**TAACGTAA**C**TCACGTGC**GACGTCGGGCCCCCTAGGTG**A-5′	Oligonucleotide containing the full-length operator but conserving only the LSE and RSE binding sites plus the 1-bp adjacent on each side (wild-type sequence: 597-613).
LSE-RSE*	55	5′-**TTGATATCGAATTCCTA**ATTGCATT**G**AGTGCACGGTTATG**CCCGGGGGATCCAC**T-3′ 3′-**AACTATAGCTTAAGGAT**TAACGTAA**C**TCACGTGCCAATAC**GGGCCCCCTAGGTG**A-5′	Oligonucleotide containing the full-length operator but conserving only the LSE plus the 1-bp adjacent on the left, and the RSE plus the 7-bp sequence adjacent on the right (wild-type sequence: 597-619).
RSE-RA	55	5′-**TTGATATCGAATTCCTAATCAAGTC**G**AGTGCACGGTTATGCTACTATAGTTTTAT-3′ 3′-**AACTATAGCTTAAGGATTAGTTCAG**C**TCACGTGCCAATACGATGATATCAAAATA-5′	Oligonucleotide containing the full-length operator but conserving only the RSE and the region corresponding to the putative RA binding site (wild-type sequence: 605-634).
LA-LSE	55	5′-GGAAAAATACGAAAAAGATTGCATT**G**GACTTGAGCTGCAGCCCGGGGGATCCAC**T-3′ 3′-CCTTTTTATGCTTTTTCTAACGTAA**C**CTGAACTCGACGTCGGGCCCCCTAGGTG**A-5′	Oligonucleotide containing the full-length operator but conserving only the region containing the putative LA binding site and the LSE (wild-type sequence: 580-605).
NS	55	5′-**TTGATATCGAATTCCTAATCAAGTCGGACTTGAGCTGCAGCCCGGGGGATCCAC**T-3′ 3′-**AACTATAGCTTAAGGATTAGTTCAGCCTGAACTCGACGTCGGGCCCCCTAGGTG**A-5′	Oligonucleotide totally modified
sLSE	55	5′-GGAAAAATACGAAAAAGA**G**TGCATT**G**AGTGCACGGTTATGCTACTATAGTTTTAT-3′ 3′-CCTTTTTATGCTTTTTCT**C**ACGTAA**C**TCACGTGCCAATACGATGATATCAAAATA-5′	Oligonucleotide containing the full-length operator with the T(598) to G change, pointed to improve the LSE by turning it into an inverted repeat of the contacted bases observed in RSE.
sRA1	55	5′-GGAAAAATACGAAAAAGATTGCATT**G**AGTGCACGGTTATGC**AC**CTATAGTTTTAT-3′ 3′-CCTTTTTATGCTTTTTCTAACGTAA**C**TCACGTGCCAATACG**TG**GATATCAAAATA-5′	Oligonucleotide containing the full-length operator with the T(621) to A and A(622) to C changes, pointed to improve the putative RA binding site by turning it into a direct repeat of the contacted bases observed in RSE.

Name	Size	Sequence	Description
sRA2	55	5′-GGAAAAATACGAAAAAGATTGCATT**G**AGTGCACGGTTATGC**C**ACTATAGTTTTAT-3′ 3′-CCTTTTTATGCTTTTTCTAACGTAA**C**TCACGTGCCAATACG**G**TGATATCAAAATA-5′	Oligonucleotide containing the full-length operator with the T(621) to C change pointed to improve the putative RA binding site.
sRA3	55	5′-GGAAAAATACGAAAAAGATTGCATT**G**AGTGCACGGTTATG**TGCA**TATAGTTTTAT-3′ 3′-CCTTTTTATGCTTTTTCTAACGTAA**C**TCACGTGCCAATAC**ACGT**ATATCAAAATA-5′	Oligonucleotide containing the full-length operator with the 5'-CTAC-3' (620-623) to 5'-TGCA-3' sequence replacement, pointed to improve the putative RA binding site by turning it into an inverted repeat of the contacted bases observed in the RSE.
sRA4	55	5′-GGAAAAATACGAAAAAGATTGCATT**G**AGTGCACGGTTATG**TG**ACTATAGTTTTAT-3′ 3′-CCTTTTTATGCTTTTTCTAACGTAA**C**TCACGTGCCAATAC**AC**TGATATCAAAATA-5′	Oligonucleotide containing the full-length operator with the C(620) to T and T(621) to G changes, pointed to improve the putative RA binding site by turning it into an inverted repeat of the contacted bases observed in the RSE.
sLA1	55	5′-GGAAAAAT**G**C**AC**AAAAGATTGCATT**G**AGTGCACGGTTATGCTACTATAGTTTTAT-3′ 3′-CCTTTTTA**C**G**TG**TTTTCTAACGTAA**C**TCACGTGCCAATACGATGATATCAAAATA-5′	Oligonucleotide containing the full-length operator with the changes: A(588) to G, G(590) to A and A(591) to C; pointed to improve the putative LA binding site by turning it into a direct repeat of the contacted bases observed in the RSE.
sLA2	55	5′-GGAAAAA**GTGC**AAAAAGATTGCATT**G**AGTGCACGGTTATGCTACTATAGTTTTAT-3′ 3′-CCTTTTT**CACG**TTTTTCTAACGTAA**C**TCACGTGCCAATACGATGATATCAAAATA-5′	Oligonucleotide containing the full-length operator with the 5'-TACG-3' (587-590) to 5'-GTGC-3' sequence replacement, pointed to improve the putative LA binding site by turning it into an inverted repeat of the contacted bases in the RSE.

[a]Coordinates of the pMV158 sequence. Nucleotides mutated in the operator sequences are in *blue*. When appears, the G:C base-pair corresponding to the two-fold symmetry axis of the SE is in *bold*.

2.4 DNA labeling

The general procedure for 5'-end ^{32}P-labeling of DNA was carried out as described by (Sambrook, Fritsch *et al.* 1989). Most of the experiments in this thesis were performed using radioactive DNA, and we proceeded in two different ways depending on whether the labeled sample was to be annealed directly or used for PCR amplification. Ratio between ssDNA and [γ^{32}P]-ATP molecules varied depending on the further application of the labeled DNA (see Table 8)**.**

Table 8 | Ratio between ssDNA and [γ-^{32}P]-ATP molecules in labeling reactions

ssDNA : [γ-^{32}P]-ATP	Application of labeled ssDNA
1 : 0.6	PCR
1 : 1	Relative binding affinity assays Cooperativity assays
1 : 1.6	OH• and DMS footprintings
1 : 3	Determination of the equilibrium dissociation constant

2.4.1. ^{32}P-labeling of oligonucleotides to be annealed

Only one of the complementary oligonucleotides was ^{32}P-labeled. In all cases, reaction volume was 50 µl, and the coding strand was labeled. For footprinting experiments, each of the two strands was separately ^{32}P-labeled and subsequently annealed with its respective non-radioactive complementary strand.

ssDNA were incubated at 70 °C for 5 min to disrupt intrastrand base pairing, and then treated with T4-PNK (New England Biolabs) in the presence of [γ^{32}P]-ATP (3000 Ci/mmol) (Perkin Elmer or Amersham) for 1 h at 37 °C, followed by incubation at 65 °C for 20 min to inactivate the enzyme. To remove the non-incorporated radioactive nucleotide, phosphorylation reactions were filtrated by liquid chromatography using the columns Micro-Spin^{TM} G25 from Amersham, following manufacturer's instructions. Co-precipitant Pellet Paint^{TM} and 10 mM $MgCl_2$ were added to improve ethanol precipitation of the oligonucleotides. ^{32}P-labeled ssDNA pellets were dissolved in Milli-Q water and annealed with its complementary strand as explained above (2.3.2). Radioactivity (cpm) was measured by a scintillation counter (Wallac 1450 MicroBeta, TriLux) before and after the precipitation step in order to calculate the specific activity (cpm/pmol) and final recovery of each DNA molecule.

2.4.2. ^{32}P-labeling of DNA fragments

To generate the 239-bp DNA fragment radiolabeled at a single 5'-end, one of the primers was ^{32}P-labeled prior to the amplification reaction. In this case, primer *Forward* 489-505 was treated with T4-PNK (New England Biolabs) in the presence of [γ^{32}P]-ATP (Amersham), and purified by gel filtration as mentioned above. Next, it was added to the PCR reaction, and the amplification was carried out as indicated before (2.2.2).

The ^{32}P-labeled PCR product was purified from 5% PAA native gels. After electrophoresis, the gel was covered with plastic Saran^{TM} wrap and exposed to X-ray film (Curix RP2 Plus, AGFA) for 2 min. The band corresponding to the labeled DNA fragment was excised from the film, and this was aligned on the gel using the fluorescent rulers attached beside it. The labeled DNA band was sliced from the gel with a scalpel-blade, and incubated in elution buffer (20 mM Tris-HCl pH 8.0, 200 mM NaCl, 1 mM EDTA) overnight at 42 °C with shaking. A 1-h second elution was carried out to increase the DNA recovery. After removal of the residual PAA fragments by filtration through "Centrifuge Tube Filter" (Spin X, COSTAR), the DNA was precipitated with isopropanol in the presence of co-precipitant Pellet Paint^{TM}. A second precipitation with ethanol was carried out to improve the purity of the DNA fragment, and the pellet was dissolved in Milli-Q water and stored at -20 °C.

3. Production of CopG proteins

Wild-type CopG protein was obtained either as reported previously (Gomis-Rüth, Solà *et al.* 1998b) or by following the protocol described below, which consists of Immobilized-Metal Affinity Chromatography (IMAC) and TAGZyme (QIAGEN) purification system. The latter protocol was also used to produce the CopG mutant proteins

3.1. Heterologous expression of *copG* genes

Blunt-ended 138-bp DNA fragments that comprise the *copG* gene variants encoding $CopG_{WT}$, $CopG_{G25E}$ or $CopG_{A30E}$ proteins were obtained by PCR amplification as specified above (2.2.1). These fragments were separately cloned into the single *Pvu*II site of the commercial TAGZyme pQE-1 expression vector (QIAGEN) (Figure 19). The DNA of the vector was linearized, dephosphorylated, and subsequently mixed with each of the three 138-bp DNA fragment variants at an insert:vector molar ratio of 2:1. T4-DNA ligase (New England Biolabs) was used for ligation in the presence of 5% polyethylene glycol and 0.1 mM ATP. The ligation reaction was allowed to proceed overnight at room temperature and then stopped by enzyme inactivation at

65 °C for 10 min. The DNA was precipitated with n-butanol, in order to remove the polyethylene glycol and most salts, and washed with ethanol. Pellets were dissolved in Milli-Q water. The *E. coli* M15 [pREP4] strain was employed for transformation with the recombinant vector. Electroporation and selection for transformants was carried out as described above (1.1 and 1.2).

Cloning of a fragment of the expected size was confirmed by PCR using Phusion® DNA polymerase (Finnzymes) and pQE-1*copG*, pQE-1*copG7* and pQE-1*copG8* DNAs as templates. Synthetic oligonucleotides *Forward* 3375-3392 and *Reverse* 248-230 (Table 4), which anneal to the vector sequence, were employed as primers, and the reaction was carried out in a final volume of 50 µl in Phusion® HF Buffer. Each component was added at the recommended concentration. The amplification consisted of 25 cycles of hybridization/elongation with temperatures of 58 °C for 20 s and 72 °C for 5 s, respectively, preceded by a denaturation step of 30 s at 98 °C. Correct orientation and sequence of the inserted fragment were confirmed by sequencing using the primer *Forward* 3375-3392, which hybridizes with vector pQE-1 upstream of the MCS (Figure 19). Sequencing was carried out by SECUGEN DNA Sequencing Facility at the CIB.

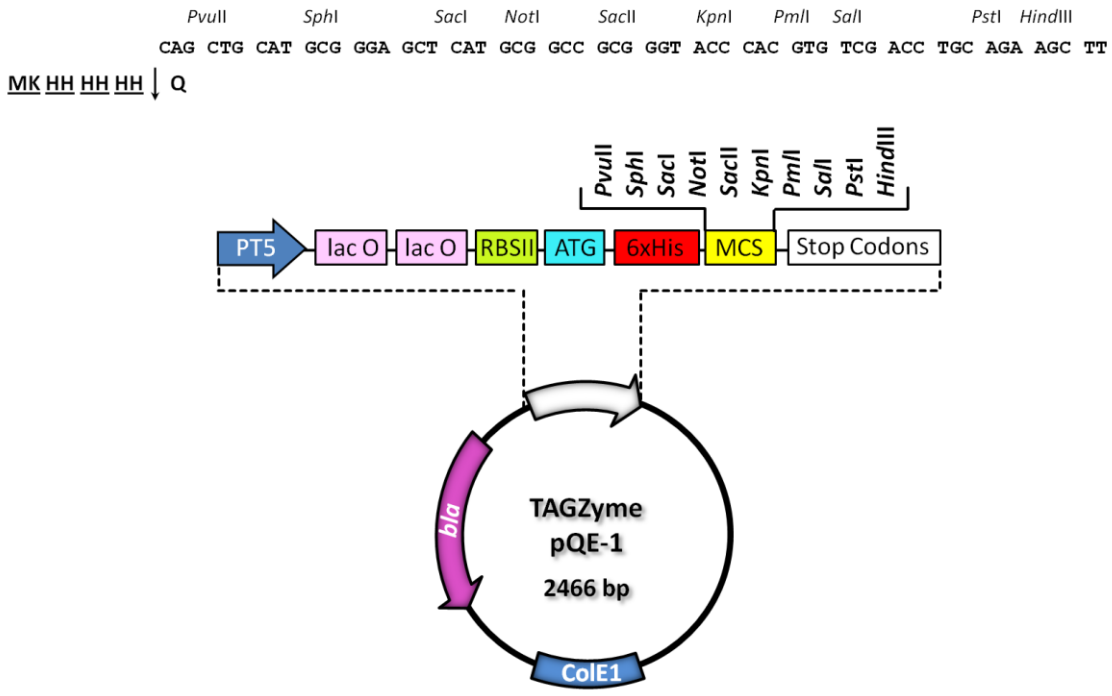


Figure 19. Functional map of the TAGZyme pQE-1 vector. *PT5*, phage T5 promoter; *lac O*, *lac* operator; *RBSII*, synthetic ribosome-binding site; *ATG*, star codon; *6xHis*, 6xHis-tag sequence; *MCS*, Multiple Cloning Site; *bla*, β-Lactamase gene which confers ampicillin resistance; *ColE1*, origin of replication.

Induced expression of the *copG* gene variants was tested in *E. coli* M15 [pREP4] transformed with the recombinant vectors. Transformants were grown in LB liquid medium as described in 1.1**.** Expression of *copG* genes cloned in pQE-1 was induced by the addition of isopropyl-β-D-thiogalactosidase (IPTG) when the culture reached an OD_{600} ~0.5. Prior to induction, a 1.5-ml volume of culture was withdrawn in order to determine the basal expression level of the *copG* gene (time 0). The induced culture was grown for 5 more hours, and 1.5-ml samples were withdrawn every hour to prepare protein crude extracts. Similar number of cells were collected from each sample by centrifugation at 4 °C, and resuspended in lysis buffer (50 mM Tris-HCl pH 6.8, 2 mM EDTA, 2% SDS, 143 mM β-mercaptoethanol, 10% glycerol). Proteins in the samples were denatured by heating at 95 °C for 3 min, and then loaded onto 16% PAA (40:1) Tris-Tricine gels. Proteins in the crude extracts were visualized by staining with Bio-Safe™ Coomassie (Bio-Rad) following the manufacturer's instructions.

3.2. Protein overproduction and purification

Large-scale induced cultures were prepared to increase the production of protein to be purified. LB liquid medium was separately inoculated with 1/50 volumes of a non-induced overnight culture of the *E. coli* M15 [pREP4]/pQE-1*copG*, *E. coli* M15 [pREP4]/pQE-1*copG7* or *E. coli* M15 [pREP4]/pQE-1*copG8* strains. Cells were grown at 37 °C for 4-5 h after IPTG-induction and then harvested by centrifugation. The cellular pellet was stored overnight at -20 °C. Once thawed, each pellet from 1 l of culture was resuspended in 40 ml of lysis buffer (20 mM sodium phosphate pH 6.0, 1.5 M NaCl). PIC (Protease Inhibitor Cocktail, SIGMA) was added at 1:1000 ratio and cells were subjected to pressure-lysis in a Constant Systems TS 1.1 KW Cell Disruptor, at 27 kpsi. Soluble proteins were separated from cellular debris by centrifugation at 40000 rpm, 4 °C for 45 min in a Beckman Ultracentrifuge and maintained at 4 °C until the IMAC was performed. Next, a volume of 2 ml of Ni-NTA (Nickel-nitrilotriacetic acid) agarose (QIAGEN, cross-linked bead structure of 6% agarose) for purification of His-tagged proteins by gravity-flow chromatography was poured in a 20 ml disposable column. The column was washed and equilibrated with equilibration buffer (20 mM sodium phosphate pH 7.4, 500 mM NaCl, 5 mM imidazole) for 10 min with gently shaking in a rotator at 4 °C. The protein-containing supernatant was then loaded in aliquots of 20 ml onto the column and kept on shaking at 4 °C for 40 min. Under these conditions of imidazole, His-tagged proteins bind to the nickel matrix. This step was repeated until all supernatant volume passed through and an additional 20 ml-volume of equilibration buffer was poured to remove any remaining non-His-tagged proteins and other impurities. Finally, the target protein was recovered from the nickel-matrix by addition of elution

buffer (20 mM sodium phosphate pH 7.4, 500 mM NaCl, 500 mM imidazole). A 5 ml-volume was poured into the column and agitated for 1 min. The protein-containing eluate was collected and buffer exchanged into 1X TAGZyme buffer (20 mM sodium phosphate pH 7.0, 150 mM NaCl) using a 3 kDa MWCO Amicon centrifugal filter. His-tagged proteins recovered in the retentate volume were then isolated throughout a high performance gel filtration, using a preparative column Superdex^{TM} 75 10/300 GL (Tricorn) connected to a FPLC^{TM} system under isocratic conditions. The column was equilibrated with 1 CV of the elution 1X TAGZyme buffer. All fractions (2 ml) generated in the gel filtration chromatogram were assessed by SDS-PAGE for protein identity. The 2-ml fractions containing the target protein were pooled together before the His-tag removal.

Elimination of the His-tag by the TAGZyme protocol was carried out following the manufacturer's instructions. The next conditions were set up to digest 2 mg of protein. A volume (15 µl) of dipeptidyl aminopeptidase I (DAPase, 0.01 U/µl) was firstly mixed with an equal volume of cysteamine-HCl (20 mM) and incubated for 5 min at room temperature (DAPase is activated by cysteamine). Next, 12 volumes (180 µl) of glutamine cyclotransferase (Qcyclase, 0.05 U/µl) were added. The enzyme mix was added to the pre-incubated (37 °C) protein solution (0.4 mg/ml) in 1X TAGZyme buffer. Digestion was performed at 37 °C for 12 h. DAPase and Qcyclase were removed by a subtractive IMAC step in Ni-NTA columns previously equilibrated with 1X TAGZyme buffer. The processed de-tagged protein was recovered in the flow-through. Subsequently, a volume of 160 µl of pyroglutamyl aminopeptidase (pGAPase, 0.025U/µl) was activated by mixing with an equal volume of cysteamine-HCl (20 mM), followed by incubation for 10 min at 37 °C. Activated pGAPase was then added to the protein-containing eluate and incubated overnight at 37 °C. Finally, the digestion reaction mixture was loaded onto Ni-NTA columns to eliminate enzymes and residual unprocessed protein. Native CopG was recovered from the flow-through and loaded onto a Superdex^{TM} 75 10/300 (Tricorn) column equilibrated with NMR buffer (20 mM sodium phosphate pH 6.0, 150 mM NaCl). Low pH is preferred for NMR experiments, since in basic solutions amide hydrogen exchange with the water of the solvent gives broader spectra signals. Fractions containing the native CopG proteins were recovered and concentrated. Proteins were visualized in Bio-Safe^{TM} Coomassie (Bio-Rad)-stained 16% PAA (40:1) Tris-Tricine gels. Concentration of the protein solutions was calculated firstly by measuring the 275 nm absorbance in a spectrophotometer and afterwards confirmed by amino acid composition quantification by the Chemistry Protein Facility at CIB. All protein samples were conserved at -80 °C. Unless otherwise stated, the protein concentrations given throughout this work refer to protomers of CopG.

Part II. EXPERIMENTAL PROCEDURES

1. Electrophoretic mobility shift assays

The electrophoretic mobility shift assay (EMSA) is a rapid and sensitive method to detect protein-nucleic acid interactions (Fried & Crothers 1981). It is based on the observation that the electrophoretic mobility of a protein-nucleic acid complex is typically lower than that of the free nucleic acid. In the classical assay, solutions of protein and nucleic acid are combined and the resulting mixtures are subjected to electrophoresis under native conditions. Most results in this thesis were obtained from EMSA experiments, probing several DNA targets mixed with a wide range of protein concentrations. The specific conditions for different experiments are detailed later, but those common for all EMSA are indicated below.

1.1. Protein-DNA complex formation

Protein-DNA complexes were basically accomplished through the incubation of a mixture of DNA and protein at the indicated concentration. Binding reactions were equilibrated in binding buffer (20 mM Tris-HCl pH 8.0, 1 mM EDTA, 5 mM DDT, 250 mM KCl) supplemented with bovine serum albumin (BSA) at 500 µg/ml. When indicated, heparin or T7-bacteriophage DNA was added as competitor. After equilibrium, and unless otherwise indicated, samples were mixed with 1X BXGE loading buffer (0.025% bromophenol blue, 0.025% xylene cyanol, 6% glycerol, 1 mM EDTA) and loaded onto native PAA gels. Percentage of PAA native gels varied (Table 9) accordingly to the size of the DNA probed.

Table 9 | PAA native gels used in EMSAs

% PAA	Experiments
5	EMSAs using the 239-bp DNA fragment
8	Competitive EMSAs using the 239-bp DNA fragment and 55-bp operator variants
10	Competitive EMSAs using 55- and 18-bp operator variants
15	EMSAs using 25-, 22-, 21- and 18-pb operators variants
	EMSAs using the 55-bp WT operator, aimed to obtain the protein footprints in individual complexes

1.2. EMSA visualization and analysis

After electrophoresis, non-radioalabeled DNAs in EMSAs were stained either with 0.5 µg/ml EtBr or with GelRed™, for 15 min with gently shaking, and imaged by the GelDoc XR documentation system (Bio-Rad).

EMSA gels with radiolabeled DNA were treated on the glass plate with a fixer solution containing 40% methanol, 10% acetic acid and 3% glycerol for 10-15 min. Thereafter, gels were well drained thoroughly and transferred carefully to a piece of Whatman® 3MM filter paper, covered with plastic Saran™ wrap and completely dried on a vacuum gel dryer (Bio-Rad). Radiolabeled DNA bands were observed by autoradiography using a storage phosphor technology, with the aid of a FLA-3000 (FujiFilm) imaging system. Bands corresponding to bound and unbound DNA were analyzed and quantified with Quantity One Software (Bio-Rad).

2. Competitive EMSAs for determining relative binding affinities

Competitive EMSA was the selected approach to determine the relative binding affinities of CopG for several DNA sequences, and was performed as described by (Ruiz-Masó, Lurz *et al.* 2007), who modified a method previously reported by (Man & Stormo 2001). Basically, two different radiolabeled DNAs were jointly incubated with varying protein concentrations and, subsequently, bound and unbound DNAs were separated on native gels. Due to the different size of the DNAs, complexes were easily differentiated.

2.1. Relative binding affinity determination

The affinity of CopG for DNA D_1 relative to DNA D_2 [$K_{rel}(D_1/D_2)$] was determined by competitive EMSAs as follows (Eq. 1):

$$K_{rel}\left(\frac{D_1}{D_2}\right) = \frac{[CopG{\cdot}D_1][D_2]}{[CopG{\cdot}D_2][D_1]} \qquad \text{(Eq. 1)}$$

where $[D_1]$ and $[D_2]$ are the concentrations of free DNA D_1 and DNA D_2, respectively. $[CopG{\cdot}D_1]$ and $[CopG{\cdot}D_2]$ are the sum of the concentrations of the complexes produced by binding of CopG to D_1 and D_2, respectively. This approximation was applied in all competitive EMSAs.

2.2. Competitive EMSAs for analyzing the role of each subsite in protein binding affinity

CopG was incubated with a mixture of two different DNAs, one of them being the 239-bp DNA fragment containing the entire wild-type operator, and the other, a 55-bp oligonucleotide from among these: WT, RSE^-, LSE^-, RA^-, LA^-, RSE^+, LSE^+, RA^+, LA^+ and NS. See Table 7 for description. DNAs were radiolabeled with [γ^{32}P]-ATP provided by Amersham. The final volume of reaction was 6 µl. CopG-DNA equilibrium binding was allowed for 10 min at room temperature and samples were loaded onto 5% PAA native gels. As the calculations were based on the amount of cpm per assay, final concentrations of DNA employed were different depending on the specific radioactivity, as shown in Table 10. The number of samples independently analyzed varied

between 17 and 33. To normalize the data for potential effects due to the different size (239 bp vs. 55 bp) and origin (PCR amplification vs. chemical synthesis) of the DNAs used in these assays, the affinities of CopG for these 55-bp operator variants relative to the 239-bp wild-type DNA fragment were divided by the ratio between the protein affinities for the 55-bp WT and 239-bp wild-type DNAs. Results are shown in *Chapter I* of *Results*.

Table 10 | DNA concentrations used in competitive EMSAs for analyzing the role of each subsite in the protein binding affinity

Labeled-DNA		Specific activity (cpm/pmol)	cpm/assay	[DNA] (nM)
239-bp	Fragment	4×10^5	2×10^4	~ 8
	WT		1×10^4	~ 4
55-bp	WT	12×10^5	1×10^4	~ 1
	RSE^-	5.5×10^5	8×10^4	~24
	LSE^-	14×10^5	4×10^4	~ 5
	RA^-	13×10^5		~ 5
	LA^-	11×10^5		~ 6
	RSE^+	9.5×10^5		~ 7
	LSE^+	10×10^5		~ 7
	RA^+	7.5×10^5		~ 9
	LA^+	9×10^5		~ 7
	NS	3×10^5		~ 22

2.3. Competitive EMSAs for analyzing the role of pairs of adjacent subsites in the cooperativity of protein binding

CopG was incubated with a mixture of two different DNAs, one of them being the 18-bp oligonucleotide o(RSE), and the other being a 55-bp oligonucleotide containing either the wild-type operator or distinct combinations of two adjacent subsites (operators: LA-LSE, LSE-RSE, LSE-RSE* and RSE-RA). DNAs were radiolabeled with [γ^{32}P]-ATP supplied either by Amersham or by Perkin Elmer. DNA concentrations depending on the specific radioactivity are summarized in Table 11. Reactions were performed in a final volume of 8 µl, in the presence of 90 ng of T7-bacteriophage DNA as competitor. Binding reaction was allowed to equilibrate for 10 min at room temperature before loading the mixtures onto 10% PAA native gels. The number of samples independently analyzed varied between 4 and 23. Non-competitive EMSAs were carried out also under these experimental conditions. Results are shown in *Chapter III* of *Results*.

Table 11 | DNA concentrations used in EMSAs for analyzing the role of pairs of adjacent subsites in the cooperativity of protein binding.

Labeled DNA		Specific activity (cpm/pmol)	cpm/assay	[DNA] (nM)
18-bp	o(RSE)	*7.5 x 10^5*	*4 x 10^4*	*~ 7*
		†10 x 10^5	*2 x 10^4*	*~ 3*
			4 x 10^4	*~ 5*
	o(RSE*)	*†12 x 10^5*	*2 x 10^4*	*~ 2*
			4 x 10^4	*~ 4*
			6 x 10^4	*~ 6*
55-bp	WT	*10 x 10^5*	*2 x 10^4*	*~ 3*
	LSE-RSE	*9 x 10^5*		*~ 3*
		†12 x 10^5		*~ 2*
	LSE-RSE*	*†14 x 10^5*		*~ 2*
	RSE-RA	*8 x 10^5*		*~ 3*
		†12 x 10^5		*~ 2*
	LA-LSE	*4 x 10^5*		*~ 6*

†[γ^{32}P]-ATP from Perkin Elmer.

2.4. Competitive EMSAs for analyzing the binding affinity of protein to the super-operators

CopG was incubated with a mixture of two different DNAs, one of them being a 239-bp DNA fragment containing the entire wild-type operator of the protein, and the other being a 55-bp oligonucleotide corresponding either to the WT operator or to each one of the super-operators (sLSE, sRA1, sRA2, sRA3, sRA4, sLA1 and sLA2). DNAs were radiolabeled with [γ^{32}P]-ATP provided by Perkin Elmer. After binding equilibrium was reached (10 min at room temperature), unbound and bound DNAs were separated in 8% PAA native gels. For these competitive EMSA, no competitor DNA was added and the final reaction volume was 8 µl. Table 12 summarizes the DNA concentrations used for these assays. The number of samples independently analyzed was 8. Using the same labeled super-operators, non-competitive EMSAs were performed in 5 µl binding reactions wherein either 90 ng of T7-bacteriophage DNA or 120 µg/ml of heparin were added as competitor. The binding reaction was allowed to reach the equilibrium at room temperature for 10 min or at 0 °C for 20 min. The latter condition was important for disclosing the complexes formed by $CopG_{G25E}$ and $CopG_{A30E}$ proteins bound to super-operators. Results are shown in *Chapter IV* of *Results*.

Table 12 | DNA concentrations used in EMSAs for analyzing the binding affinity of protein for super-operators

Labeled-DNA		Specific activity (cpm/pmol)	cpm/ assay	Competitive EMSA [DNA](nM)	Non-competitive EMSA [DNA] (nM)
239-bp	Fragment WT	4×10^5	1×10^4	~ 3	-
55-bp	WT	6×10^5	2×10^4	~ 4	-
	sRA2		4×10^4	-	~ 13
	sRA3				
	sLSE	7.5×10^5	2×10^4	~ 3	-
			4×10^4	-	~ 11
	sRA1	8×10^5	2×10^4	~ 3	-
			4×10^4	-	~ 10
	sRA4	5.5×10^5	2×10^4	~ 5	-
			4×10^4	-	~ 14
	sLA1	6.5×10^5	2×10^4	~ 4	-
			4×10^4	-	~ 12
	sLA2	8.5×10^5	2×10^4	~ 3	-
			4×10^4	-	~ 9

2.5. Competitive EMSAs for analyzing the binding of protein to super-operators consisting only of two subsites

Competitive EMSAs using operator variants consisting exclusively of two subsites were also performed mixing each one of them (see Table 7) with the 55-bp oligonucleotide having the wild-type operator and several protein concentrations. T7-bacteriophage DNA (90 ng) was added as competitor and the binding reactions (8 µl final volume) were allowed to reach the equilibrium at room temperature for 10 min. Samples were loaded onto 10% PAA native gels. Non-competitive EMSAs were carried out under the same experimental conditions but complexes and free DNA were separated either on 10% or on 15% PAA native gels. All DNAs were radiolabeled with [γ^{32}P]-ATP supplied by Perkin Elmer. Table 13 shows the DNA amounts and concentrations employed. Results are shown in *Chapter IV* of *Results.*

Table 13 | DNA concentrations used in EMSAs for analyzing the binding of protein to super-operators consisting only of two subsites

Labeled-DNA		Specific activity (cpm/pmol)	cpm/assay	[DNA] (nM)
55-bp	WT	*4.5 x10^5*	*1 x 10^4*	~ 3
21-bp	o(LSE-RSE)	*6 x10^5*	*2 x 10^4*	~ 4
	o(sLSE-RSE)			
22-bp	o(sLSE+1-RSE)			
25-bp	o(RSE-RA)	*5 x10^5*		~ 5
	o(LA-LSE)			
	o(RSE-1-sRA3)			
	o(RSE-sRA1)	*2.8 x10^5*		~ 9
	o(RSE+1-sRA1)	*3.5 x10^5*		~ 7
	o(RSE-sRA3)	*4.5 x10^5*		~ 6
	o(sLA1-LSE)			
	o(sLA1-1-LSE)	*5.5 x10^5*		~ 5

3. EMSAs for analyzing protein binding cooperativity

For analyzing the cooperativity of CopG binding to its operator DNA, 55-bp operator variants containing only two adjacent wild-type subsites (LSE-RSE, LSE-RSE*, RSE-RA and LA-LSE) as well as the short oligonucleotides consisting exclusively of two subsites [21-bp o(LSE-RSE), 25-bp o(RSE-RA) and 25-bp o(LA-LSE)] were employed in EMSAs.

Table 14 | DNA concentrations used in EMSAs for cooperativity analysis

Labeled DNA		Specific activity (cpm/pmol)	cpm/assay	[DNA] (nM)
21-bp	o(LSE-RSE)	*6 x 10^5* *6.5 x 10^5*	*~ 1.2 x 10^4* *~ 1.3 x 10^4*	2.5
25-bp	o(RSE-RA)	*5 x 10^5*	*~ 1 x10^4*	
	o(LA-LSE)			
55-bp	LSE-RSE	*†9 x 10^5* *12 x 10^5*	*2 x 10^4*	~ 3 ~ 2
	LSE-RSE*	*15 x 10^5*		~ 2
	RSE-RA	*†8 x 10^5* *12 x 10^5*		~ 3 ~ 2
	LA-LSE	*†4 x 10^5*		~ 6

†Samples labeled with [γ^{32}P]-ATP from Amersham.

All of these operator variants were mixed and allowed to equilibrate (for 10 min at room temperature) with increasing amounts of protein in the presence of T7-bacteriophage DNA (60-90 ng) as competitor. The final volume of reaction was 8 µl and all samples were loaded onto 15% PAA native gels. The number of samples independently analyzed varied between 8 and 14. Table 14 shows the DNA concentrations used for these experiments. Results are shown in *Chapter III* of *Results*.

3.1. Cooperativity analysis

Data obtained from EMSA were analyzed according to the model described by (Senear & Brenowitz 1991). The model for a system composed of two interacting DNA binding sites and a single protein ligand is shown in Table 15. The relative probability of each of the configurations in the table is given by:

$$f_s = \frac{e^{-\Delta G_S/RT} L_j}{\sum_{sj} e^{-\Delta G_S/RT} L_j} \quad \text{(Eq. 2)}$$

where $-\Delta G_s$ is the sum of free energy contributions for configuration *s*, *R* is the gas constant, *T* is the absolute temperature, *L* is the concentration of free protein ligand, and *j* is the number of ligands bound to configuration *s*.

The (binding) equations for the Θ_i, which means the fraction of DNA molecules with exactly *i* ligands bound, are obtained by summing the relative probabilities of the appropriate configurations. For a two-site system (Table 15), this yields:

$$\Theta_0 = f_1 \quad \text{(Eq. 3a)}$$

$$\Theta_1 = f_2 + f_3 \quad \text{(Eq. 3b)}$$

$$\Theta_2 = f_4 \quad \text{(Eq. 3c)}$$

These equations are greatly simplified by making familiar substitutions $k = e^{-\Delta G/RT}$ to give:

$$\Theta_0 = 1/Z \quad \text{(Eq. 4a)}$$

$$\Theta_1 = (k_1 + k_2) \cdot L/Z \quad \text{(Eq. 4b)}$$

$$\Theta_2 = (k_1 k_2 k_{12}) \cdot L^2/Z \quad \text{(Eq. 4c)}$$

where *Z* is the binding polynomial (Wyman 1964) equal to 1 + $(k_1+k_2) \cdot L + (k_1k_2k_{12}) \cdot L^2$; k_1 and k_2 are the microscopic equilibrium association constants for intrinsic binding sites 1 and 2; k_{12} is the constant describing cooperative interactions when both sites are liganded.

Cooperativity and its potential involvement in a regulatory mechanism are assessed based on k_{12} (or the corresponding ΔG). However, since the three microscopic equilibrium constants, k_1, k_2, and k_{12}, appear in only two combinations, they can be replaced by two macroscopic equilibrium constants defined by $K_1 = k_1 + k_2$ and $K_2 = k_1k_2k_{12}$:

$$\Theta_0 = 1/Z \qquad \text{(Eq. 4a')}$$

$$\Theta_1 = K_1 \cdot L/Z \qquad \text{(Eq. 4b')}$$

$$\Theta_2 = K_2 \cdot L^2/Z \qquad \text{(Eq. 4c')}$$

Only these macroscopic product constants, K_1 and K_2, can be determined from a single mobility-shift experiment, and the cooperative interactions are inferred by comparing them. Considering a two-site system, the binding is cooperative if $K_2 > K_1^2/4$.

Table 15 | Configurations and associated free energy states for protein binding to DNA

Two sites (general case)[a]					
Species	Binding Configurations			Free energy contributions	Total free energy
	Site 1		Site 2		
1				Reference state	ΔG_{s1}
2			L	ΔG_1	ΔG_{s2}
3	L			ΔG_2	ΔG_{s3}
4	L	$\longleftrightarrow$	L	$\Delta G_1 + \Delta G_2 + \Delta G_{12}$	ΔG_{s4}

[a]Binding sites are denoted by L if liganded. Cooperative interactions are denoted by $\longleftrightarrow$. The total Gibbs free energy of each configuration (ΔG_s) relative to the unliganded reference state is given as the sum of contributions from three energy changes (column 3). ΔG_i (i = 1 or 2) are the intrinsic free energy changes for binding to the individual operator sites.

4. Determination of the equilibrium dissociation constant for binding of CopG proteins to the RSE

EMSAs were performed employing the radiolabeled 18-bp oligonucleotide comprising exclusively the RSE (Table 7) mixed with increasing amounts of protein. DNA was labeled with [γ^{32}P]-ATP supplied by Amersham. Binding reactions were performed in a final volume of 8 µl, and held for 30 min at 0 °C. Next, samples were mixed with a non-dye loading buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA, 50% glycerol) and loaded onto 30 V-running 15% PAA native gels to

avoid disaggregation of unstable complexes. Electrophoresis was carried out at 0 °C. DNA concentrations used are indicated in Table 16.

Table 16 | DNA concentrations used in EMSAs to calculate the K_D of CopG-RSE binding

Specific activity (cpm/pmol)	cpm/assay	pmol/assay	[DNA] nM	Protein assayed
8.6×10^5	$\sim 3.5 \times 10^2$	4×10^{-4}	0.05	$CopG_{WT}$
	$\sim 3.5 \times 10^3$	4×10^{-3}	0.5	$CopG_{G25E}$
	$\sim 2.8 \times 10^4$	3.2×10^{-2}	4	$CopG_{A30E}$

4.1. K_D calculation

The affinity was determined as the inversed of the dissociation constant (K_D) of the binding equilibrium, which under the experimental conditions employed (large molar excess of protein relative to DNA) represents the protein concentration required to have a 50% of the DNA bound. This parameter was obtained by non-linear regression fitting of the experimental data to the following expression (Eq. 5) using the SigmaPlot ® 10.0 software:

$$Y = \frac{B_{max}\, X}{K_d + X} \qquad \text{(Eq. 5)}$$

where B_{max} denotes the maximum fraction of DNA bound that could be estimated from data, X is the concentration of protein and, Y corresponds to the fraction of DNA bound at a given protein-concentration (X). Results are shown in *Chapter III* of *Results*.

5. Footprinting experiments

Footprinting is a widely used method, introduced by (Galas & Schmitz 1978), for studying the interactions between proteins or small molecules and nucleic acids. The principle of this method is based on the fact that, a protein that is bound to a specific DNA sequence, shields the DNA duplex from attack through an enzyme or chemical reagent, which is otherwise free to react with the unbound portions of the DNA molecule. Therefore, the binding sites of proteins are revealed as the DNA nucleotides that are not cleaved by the chemical or enzyme. In practice, the cleavage pattern of a protein-DNA complex is compared to that of a DNA molecule treated with the cleaving reagent in the absence of protein. This control sample provides the baseline pattern of cleavage for that DNA sequence. The cleavage pattern is visualized as a ladder of bands after denaturing gel electrophoresis of singly end-labeled DNA. Thus, the missing (or diminished)

bands in the lane that is derived from the cleavage of the protein-DNA complex constitute the "footprint".

For this thesis, two different techniques were followed to analyze the protein-DNA interactions by footprinting. Protein-DNA complexes were treated with hydroxyl radical (OH•) or dimethyl sulfate (DMS). For both approximations, protein-DNA complexes were accomplished as described (Part II, 1.1) in a final volume of 50 µl. The different complexes formed, either treated with OH• or before being treated with DMS, were separated and eluted from PAA native gels and then analyzed by sequencing on denaturing gels. Oligonucleotides used in footprinting experiments were all labeled with [γ^{32}P]-ATP supplied by Perkin Elmer. In order to improve the visualization of the lowly-populated complexes and to have sufficient intensity of radioactivity to complete the procedure, we adjusted the amount of radiolabeled DNA to 4 x 10^6 cpm per footprinting assay by varying the initial DNA concentration inversely to the specific radioactivity obtained for each oligonucleotide (Table 17). Therein the binding and PAA native gels-electrophoresis conditions are also indicated.

Table 17 | Initial DNA concentrations and general conditions used to obtain the OH• and DMS footprints of protein-DNA complexes

DNA	Labeled strand[a]	Specific activity (cpm/pmol)	cpm/ assay	[DNA] (nM)	Protein assayed	Binding & run of native gels
WT	A	*Ranging from 6 x 10^5 to 1 x 10^6*	*4 x 10^6*	~ 7 - 13	$CopG_{WT}$ $CopG_{G25E}$ $CopG_{A30E}$	0 °C + Heparin (120 µg/ml)
	B	*Ranging from: 4.5 x 10^5 to 1 x 10^6*		~ 7 -18		
sLSE	A	*7.5 x 10^5*		~ 11	$CopG_{G25E}$	
sRA3	A	*6 x 10^5*		~ 13		

[a] Strands *A* and *B* correspond to *coding* and *non-coding* strands, respectively

5.1. Hydroxyl radical footprinting

Three compounds were combined in the footprinting reaction mixture to produce OH•: $Fe(EDTA)]^{2-}$, hydrogen peroxide (H_2O_2) and sodium ascorbate. The Fenton reaction, by which the hydroxyl radical is generated, is shown in Eq. 6.

$$[Fe(EDTA)]^{2-} + H_2O_2 \xrightarrow[\text{Ascorbate}]{} [Fe(EDTA)]^{-} + OH^{-} + OH\bullet \qquad \text{(Eq.6)}$$

In this reaction an electron from iron(II)-EDTA serves to reduce and break the O-O bond in hydrogen peroxide, giving as products iron(III)-EDTA, the hydroxide ion, and the neutral hydroxyl radical. Sodium ascorbate is present to reduce the iron(III) product to iron(II), thereby establishing a catalytic cycle and permitting low (micromolar) concentrations of iron(II)-EDTA to be effective in cleaving DNA (Dixon, Hayes *et al.* 1991). A consequence of this scheme is that the concentrations of the three chemical species may be varied to optimize the generation of OH• under different solution conditions.

OH• footprinting reactions were all performed based on the protocol published by (Tullius & Dombroski 1986). The reacting solutions were 10X concentrated (except the sodium ascorbate) relative to those reported (Tullius & Dombroski 1986). With the exception of the 8 mM EDTA (Aldrich), which could be stable for months when stored at 4 °C, the other solutions, namely 4 mM iron(II) [$(NH_4)_2Fe(SO_4)_2 \cdot 6H_2O$] (ALDRICH), 20 mM sodium ascorbate (SIGMA) and 6% H_2O_2 (Merck), as well as a stock solution of $Fe(EDTA)^{2-}$ (prepared by mixing equal volumes of 4 mM iron(II) and 8 mM EDTA), were prepared immediately before use.

5.1.1. OH• treatment of free DNA and protein-DNA complexes

After allowing binding to reach the equilibrium, footprinting reaction was initiated by placing carefully 3 µl of each of the following three solutions as separate drops on the inner wall of the Eppendorf tube: $Fe(EDTA)^{2-}$ (2 mM Fe(II)/4 mM EDTA), 6% H_2O_2 and 20 mM sodium ascorbate. Using a pipette tip, drops were mixed and then quickly added to the protein-DNA mixture. The OH• cleavage was allowed to proceed for 7 min at 0 °C, and then quenched by adding thiourea and EDTA at the final concentrations of 9.5 mM and 1.7 mM, respectively. The treated samples containing the protein-DNA complexes were then loaded onto 30 V-running 15% PAA native gels, to avoid disaggregation of unstable complexes, and electrophoresed at 0 °C at 120 V until optimal separation of complexes was achieved (4-5 h).

5.1.2. Recovery of OH• -treated samples from native PAA gels

After electrophoresis was finished, gel was dismounted and, for excising of bands, it was covered with plastic Saran™ wrap and exposed to X-ray film (Curix RP2 Plus, AGFA) for 10 min. The film was developed and the bands corresponding to the complexes were outlined on the film. This film was superposed on the gel, adjusting the autoradiography with the fluorescent rules attached besides it. The bands were excised from the gel with a scalpel-blade and thereafter, DNA was recovered by elution in buffer (20 mM Tris-HCl pH 8.0, 200 mM NaCl, 1 mM EDTA) overnight at 42 °C with shaking. A 1-h second elution was carried out to increase the DNA

recovery. To eliminate residual PAA, samples were filtrated on "Centrifuge Tube Filter" (Spin X, COSTAR). Co-precipitant Pellet Paint™ and 0.1 volume of 3 M sodium acetate were added before ethanol precipitation. Dry pellets were dissolved in 0.5X sequencing gel-loading buffer (80% deionized formamide, 10 mM NaOH, 1 mM EDTA, 0.1% bromophenol blue, 0.1% xylene cyanol), and loaded onto a sequencing gel (section 5.3). The Maxam and Gilbert reactions were also loaded as controls of sequence (section 6).

5.2. Dimethyl-sulfate footprinting

Unlike for the OH• cleavage procedure, the DMS treatment was carried out after the free and bound DNA samples were separated in PAA native gels. Once the binding reaction reached the equilibrium, the samples were loaded onto 30 V-running 15% PAA native gels, to avoid disaggregation of unstable complexes, and electrophoresed at 0 °C at 120 V until optimal separation of complexes was achieved (4-5 h).

After electrophoresis was finished, the gel was covered with plastic Saran™ wrap and exposed to X-ray film (Curix RP2 Plus, AGFA) for 10 min. The film was developed and the bands corresponding to both unbound and bound DNA were outlined on the film. This film was superposed on the gel, adjusting the autoradiography with the fluorescent rules attached besides it. The bands were excise from the gel with a scalpel-blade and the PAA-pieces were put separately within Eppendorf tubes.

5.2.1. DMS treatment of free DNA and protein-DNA complexes

In the Eppendorf tube, the samples were equilibrated in 10 mM Tris-HCl pH 8.0, and after the solution was discarded, an equivalent volume of 0.2 % DMS (v/v) was added. The methylation reaction was allowed to proceed for 5-7 min at room temperature and then, the solution was completely removed. Residual methylation was quenched by adding a stop solution (150 mM Tris-HCl pH 7.3, 5 mM EDTA, 0.5 M β-mercaptoethanol) and incubating for 10 min at 37 °C.

5.2.2. Recovery of DMS-treated samples

Methylated DNA was recovered from the PAA-residual pieces by elution in buffer (0.5 M ammonium acetate pH 7.5, 0.1% SDS, 1 mM EDTA, 0.1 M β-mercaptoethanol) overnight at 42 °C with shaking. A 1-h second elution was performed to increase the DNA recovery and rubbish PAA was eliminated by filtration on "Centrifuge Tube Filter" (Spin X, COSTAR). Filtered supernatants were subjected to two sequential organic extractions, the former with a mixture of phenol-chloroform (1:1) and the latter with chloroform-isoamyl alcohol (24:1). DNAs were

recovered by precipitation with ethanol in the presence of co-precipitant Pellet Paint™ and 0.1 volume of 3 M sodium acetate. Once the pellets were washed and completely dried, DNAs were subjected to standard cleavage reactions with piperidine. Hereafter, the procedure was the same as described for sequencing reactions (section 6). The treated samples were loaded onto denaturing gels along with the Maxam and Gilbert reactions used as controls of sequence.

5.3. Sequencing gels for the analysis of footprinting experiments

To equalize the intensity of radioactivity per lane in the sequencing gel, DNA pellets arising from different specific complexes were counted, and each pellet was accordingly dissolved in the appropriate volume of loading buffer so as to obtain about 10000 cpm/µl. Samples were heated at 95 °C for 3 min and loaded onto pre-run denaturing gels. See Part I, 2.3.1 and Table 4 for details of denaturing gels. Samples were electrophoresed at constant power (45-50 W) for sufficient time to resolve the region of interest.

After electrophoresis, gels were allowed to cool prior to dismounting them. Siliconized glass plate was carefully removed from the gel assembly and immediately, a large piece of Whatman™ 3MM paper filter was employed to cover the exposed gel. Gel was peeled off along the paper and covered with plastic Saran™ wrap, then vacuum-dried, and exposed to a phosphor screen (IP FujiFilm). Bands were detected by a phosphorimager (FLA-3000, Fuji-Film) scanner and analyzed by the Quantity One Software (Bio-Rad).

6. Sequencing reactions (Maxam & Gilbert)

Sequencing reactions of DNA were performed basically as described by (Maxam & Gilbert 1980). Cleavage at purine residues (A+G) for every labeled strand was used as control of sequence in footprinting denaturing gels, and was carried out as follows.

Single-end-labeled DNA (4 x 10^4 cpm) was incubated in the presence of 65 mM formic acid and 1 µg tRNA for 1 min at 65 °C. Next, 250 mM sodium acetate pH 7.0 and co-precipitant Pellet Paint™ were added before ethanol precipitation. Pellets were completely vacuum-dried. Subsequently, DNA was dissolved in 50 µl of 1 M piperidine (Merck). Piperidine is used to cleave the sugar-phosphate chain of the DNA at the sites of chemical modification induced by the formic acid in the first step. The reaction with piperidine was run for 10 min at 100 °C. Next, samples were chilled on ice and transferred to a new Eppendorf tube. DNA was extracted by adding 1 ml of 1-butanol and precipitated by mixing with 500 µl 1-butanol and 50 µl of 1% SDS. The DNA was finally washed with ethanol and dissolved in 0.5X sequencing gel-loading buffer

(80% deionized formamide, 10 mM NaOH, 1 mM EDTA, 0.1% bromophenol blue, 0.1% xylene cyanol). Samples were stored at -20 °C.

7. Analytical ultracentrifugation experiments

Analytical ultracentrifugation experiments were carried out by the Analytical Ultracentrifugation Facility at CIB. This approach is a useful method for detailed characterization of the stoichiometry of macromolecular interactions in solution. Therefore, sedimentation equilibrium and velocity experiments were performed to study quantitatively the formation of macromolecular complexes by binding of CopG to its target DNA.

Assays were carried out using either Optima XL-A or Optima XL-I analytical ultracentrifuge (Beckman-Coulter) equipped with UV-visible absorbance detection systems. An An-Ti60 rotor and standard (12-mm optical path) double sector centerpieces were used. Absorbance scans of samples were carried out at 257 nm.

The sedimentation coefficient distributions were determined by direct linear least-squares boundary fitting of the sedimentation velocity profiles using the software SEDFIT (version 12.0d) (Schuck 2000) and HETEROANALYSIS 1.1.44.

7.1. Sedimentation equilibrium

The WT 55-bp oligonucleotide containing the full-length wild-type operator (300 nM) was mixed separately with different protein concentrations. Binding reaction was performed at room temperature in a final volume of 90 µl in buffer containing 20 mM Tris-HCl pH 8.0, 0.25 mM EDTA, 0.013 mM DTT, 0.6% ethylene glycol, 115 mM NaCl, and 500 µg/ml heparin. A sample (4 µl) of each reaction was loaded onto a 15% PAA native gel. DNA alone was used as control under the same experimental conditions. For sedimentation analysis, samples were subjected to centrifugation at 9500 rpm, 20 °C, for 18-24 h.

Buoyant molecular weights (bM_W) were obtained by fitting the equation describing the radial concentration distribution of an ideal solute at sedimentation equilibrium to the experimental data using the EQASSOC program (Minton 1994). The number of CopG molecules bound per DNA molecule (*CopG molecules/DNA molecule*) was calculated as follows:

$$\text{CopG molecules/DNA molecule} = \frac{bM_{w,CopG-DNA} - bM_{w,DNA}}{bM_{w,CopG}} \qquad \text{(Eq. 7)}$$

Where $bM_{w,CopG\text{-}DNA}$ refers to the average buoyant masses of the samples containing protein; $bM_{w,DNA}$, the buoyant mass of the DNA molecule alone, and $bM_{w,CopG}$ the buoyant mass of CopG molecule. Since total protein concentration $[CopG_{Total}]$ and DNA concentration *[DNA]* were known, the amount of free protein *[P]* was calculated as follows:

$$\text{CopG molecules/DNA molecule} \times [DNA] = [\text{CopG bound to DNA}] \quad \text{(Eq. 8)}$$

$$[P] = [CopG_{Total}] - [\text{CopG bound to DNA}] \quad \text{(Eq. 9)}$$

Experimental data arising from sedimentation equilibrium analysis were fitted to Eq. 10 using the non-linear regression analysis of SigmaPlot® (version 10.0).

$$y = \frac{n\,[P]^c}{K^c + [P]^c} \quad \text{(Eq. 10)}$$

Where *y* means, in this case, the average number of CopG molecules bound per DNA molecule; *n* is the total number of equivalent sites suitable for binding; *K*, represents the concentration of free protein that gives half-maximal binding; and the parameter *c* is known as the Hill constant or coefficient ($1 \leq c \leq n$). When c = 1, there is no cooperativity between the sites, and if, on the other hand, c = n, we would have completely cooperative binding (Rippe 1997).

7.2. Sedimentation velocity

The 55-bp oligonucleotide bearing the entire wild-type operator (300 nM) was mixed with various amounts of protein in buffer containing 20 mM Tris-HCl pH 8.0, 0.25 mM EDTA, 0.013 mM DTT, 0.6% ethylene glycol, 125 mM NaCl, and 500 µg/ml of heparin. The final volume of reaction was 300 µl. After allowing binding to reach the equilibrium at room temperature, samples were ultracentrifuged for about 2 h at 48000 rpm and 20 °C. The sedimentation coefficients were determined and analyzed using the software SEDFIT version 12.0d (Schuck 2000) and SVEDBERG (Philo 1997).

8. Circular dichroism measurements

Circular dichroism (CD) spectra were acquired in a Jasco-810 spectropolarimeter fitted with a peltier temperature control accessory. Far-UV spectra were recorded in 0.1- to 1-cm optical path length quartz cells over a wavelength range from 200 to 280 nm at a temperature of 20 °C. Each CD spectrum was taken at a scanning speed of 10 nm/min and 1-nm spectral bandwidth. CD spectrum of the buffer alone was subtracted from experimental spectra. Proteins were diluted in phosphate buffered saline (PBS), at pH 7.4.

8.1. Thermal denaturation

The protein unfolding process was monitored by the loss of the ellipticity signal at 220 nm during increase of temperature from 20 to 99 °C. Denaturation curves were obtained with a slope of 1 °C/min. Proteins spectra were recorded (as described above) both before and after thermal denaturation. CD data of denaturation curves were fitted to the following equations:

$$K_F = exp\left(\frac{\Delta H}{RT}\right)\left(\frac{T}{T_m} - 1\right) - \ln P_T \quad \text{(Eq. 11)}$$

$$FF = \frac{(4P_T K_F + 1) - (8P_T K_F + 1)^{1/2}}{4PK} \quad \text{(Eq. 12)}$$

$$\theta_T = (\theta_F - \theta_U)\, FF + \theta_U \quad \text{(Eq.13)}$$

where Θ_T is the data CD signal, Θ_F *and* Θ_U the component of the total CD corresponding to the *F*olded and *U*nfolded fractions respectively; *FF*, *F*olded *F*raction; K_F, the equilibrium constant for the unfolding process; P_T is the total concentration of protein; *T*, temperature in *°K*; T_m, melting temperature, *R*, gas constant (1.987 cal/mol); *ΔH*, enthalpy change.

9. Cloning of the super-operators in pC194 and *in vivo* assays

For cloning, *Hin*dIII-compatible 5'-AGCT-3' overhangs were added by chemical synthesis at each of the 5'-ends of the 55-bp oligonucleotides sLSE and sRA3 (Table 7, Figure 20). These super-operators were cloned into the single *Hin*dIII site of plasmid pC194 as follows.

DNA of plasmid pC194, previously linearized by digestion with *Hin*dIII enzyme (New England BioLabs), was mixed separately with each of the super-operators at a molar ratio of 10:1 insert:vector. The sLSE and sRA3 oligonucleotides were phosphorylated by adding T4-PNK (New England BioLabs) to the mixture, and the reaction of phosphorylation was kept for 1 h at 37 °C. Next, samples were heated at 65 °C for 20 min to inactivate the enzyme prior to addition of T4-DNA Ligase (New England BioLabs). Reactions of ligation were incubated overnight at room temperature, and subsequently heated at 65 °C for 20 min for enzyme inactivation. The DNA was precipitated and analyzed by migration on 1% agarose gels. Since no *Hin*dIII site will regenerate after insertion of oligonucleotides (Figure 20), selective linearization of non-recombinant religated pC194 DNA was carried out by incubating the ligation mixture with *Hin*dIII for 2 h at 37 °C. Next, the DNA was treated with phenol:chloroform and ethanol-precipitated.

Competent pneumococcal cells were transformed (with the pC194-based recombinant vectors) and screened for transformants, as previously described (Part I, 1.2). Pneumococcal DNA plasmids were purified as described (del Solar, Puyet *et al.* 1987) and the insertion of super-operators was confirmed by the no-digestion with *Hin*dIII enzyme. Since super-operators have a single *Apa*LI site in the region corresponding to the RSE, another way to confirm the successful insertion of super-operators into pC194 was to digest the recombinant vectors extracted from transformants with *Apa*LI enzyme (Figure 20). Samples from both reactions of digestion were analyzed in 1% agarose gels. Orientation of the insertions was validated by sequencing using the primer *Reverse* 2534-2518, which hybridized with the pC194 DNA downstream of the insertion site. Sequencing was carried out by SECUGEN DNA Sequencing Facility at CIB.

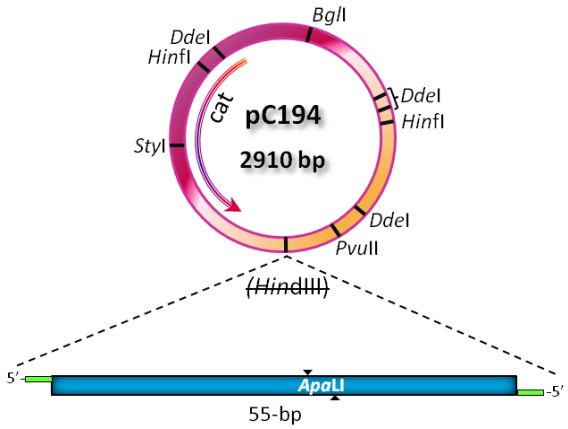


Figure 20. Plasmid pC194 and the site of insertion of super-operators. The 55-bp oligonucleotides were provided with 4-bp 5'-end overhangs (*green*) to hybridize with cohesive ends of previously *Hin*dIII-linearized pC194 DNA. After insertion, the *Hin*dIII site was not regenerated. Site *Apa*LI of insert is also shown.

Thereafter, pneumococcal cells harboring the pLS1 plasmid were transformed with the recombinant pC194 plasmids and selected for Cm^R as described in Part I, 1.2. Several clones were grown and total DNA (crude extracts) was prepared as indicated (Part I, 2.1.1). The DNA was electrophoresed on 1% agarose gels, in 1X TAE buffer (Tris base 40 mM pH 8.1, acetic acid 20 mM, EDTA 2 mM) for 16-18 h at slow voltage (2 V/cm). Chromosomal and plasmidic DNAs were visualized by staining with GelRed™ (15 min with gently shaking) and imaged by the GelDoc XR documentation system (Bio-Rad). Bands were analyzed and quantified by using the Quantity One Software (Bio-Rad). Data were normalized for the average copy number (*N*-value) of pLS1 (22 ± 2) in the wild-type condition.

GUARANTEE OF QUALITY

All chemical products were supplied by prestigious trading providers, with the higher purity and quality, and used following the manufacturer`s recommendations. Solutions and experimentation procedures were performed under the best conditions, sterilized when necessary. Water employed was purified through a Millipore Milli-Q system and then sterilized as was the non-disposable glassware.

Results

CHAPTER I

Binding of repressor CopG to its operator DNA

Transcriptional repressor CopG has been shown to be a dimer in solution at concentrations ranging from 10 to 800 µM, with no detectable monomers or association states of higher order than dimers (Acebo, García de Lacoba *et al.* 1998). Therefore, a CopG dimer is thought to be the active repressor form. Previous OH• footprinting analysis showed that contacts of the protein with its DNA operator span about 50 bp on the same face of the double helix. Moreover, mobility shift assays showed the appearance of four well-defined complexes, named CI, CII, CIII and CIV, with a cooperative binding pattern (Figure 21). These complexes are likely to contain one, two, three and four dimers bound, respectively. Based on these observations, a series of experiments were designed to prove that our working model of the P_{cr}-promoter repressosome, a complex consisting of four CopG dimers bound to the DNA operator, is reached by the consecutive binding of dimers that is stabilized by protein-protein interactions.

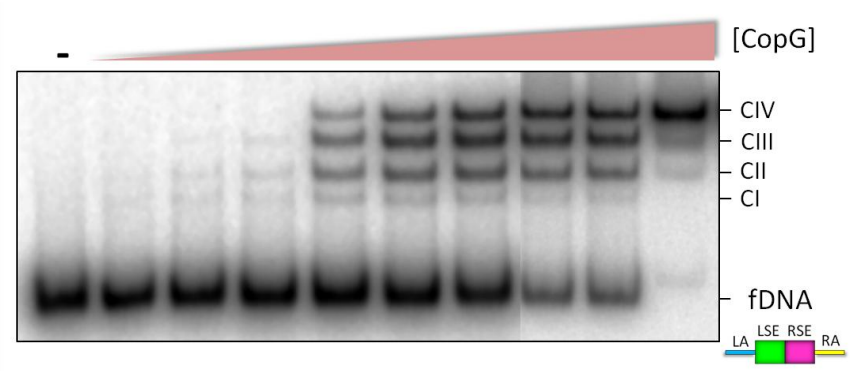


Figure 21. Complex formation of CopG with its operator DNA displaying a typical cooperative pattern. Successive binding of the protein results in generation of four complexes (CI-CIV). Bound and unbound (fDNA) DNA forms are indicated. Scheme: c*olored boxes with filled lines* represent the subsites (*blue* for LA, *green* for LSE, *magenta* for RSE, and *yellow* for RA).

1. Detection of CopG-DNA complex formation by sedimentation velocity

The binding of CopG to the WT oligonucleotide comprising the full-length wild-type operator (Table 7) was studied by sedimentation velocity analytical ultracentrifugation, employing a fixed amount of DNA (0.3 µM) in the absence or presence of varying concentrations of CopG (2.4 and 4.8 µM). Results are summarized in Table 18. The species III was separated in species IIIa and IIIb because, despite sedimenting similarly, they came from different types of sample.

Figure 22 shows the distribution of sedimentation coefficients, corrected for water at 20 °C ($S_{20,W}$), of the three samples assayed. The DNA alone sedimented mainly as a single species with an *s*-value of 3.6 S (species II in Figure 22), and the presence of two minor peaks, with values of 1.8 and 5.3 S (species I and IIIa, respectively) is likely due to degradation or random aggregation of DNA molecules. In the mixture of protein and DNA at a molar ratio of 8:1 (2.4 µM of CopG),

two separated peaks appeared; the slower one corresponding to the naked DNA, and the faster species, with an *s*-value of 5.4 S, representing the nucleoprotein complex (species IIIb). This species accounted for about 80% of the absorbance, showing that, at this concentration of CopG (2.4 μM), the majority of the DNA molecules are interacting with protein. Increasing the CopG concentration to a protein-to-DNA molar ratio of 16:1 (4.8 μM of CopG) resulted in a higher proportion of the nucleoprotein complex, about 90% of the total absorbance. The sedimentation coefficient distribution showed that it sedimented slightly faster, with an *s*-value of 5.7 S, than the nucleoprotein complex obtained at the lower protein concentration (Figure 22, Table 18). On the other hand, albeit the peak sedimenting with an *s*-value of 2.25 S appears between species I and II from other samples, we assumed it to be the unbound DNA (species II) of sample with 4.8 μM of CopG (Table 18, Figure 22). The higher concentration of protein could be increasing the viscosity of the environment, thus preventing its correct sedimentation.

Although this technique does not allow resolution of nucleoprotein complexes of different order hence yielding average *s*-values, the results obtained suggested the formation of higher-order complexes as the protein concentration was increased. The peak observed at the higher protein concentration (4.8 μM) mainly corresponded to the specific complex formed by four CopG dimers bound to the operator DNA, as demonstrated by sedimentation equilibrium and parallel EMSA analysis of replicates of these samples (Figure 23, see below). The EMSA analysis also corroborated the sedimentation velocity data, showing that the unbound DNA was less than 35% or 10% at total protein concentration of 2.4 μM or 4.8 μM, respectively.

Table 18 |Summary of sedimentation velocity centrifugation data for CopG combined with the WT operator

	Species									
	I		**II**		**IIIa**		**IIIb**		**IV**	
CopG:DNA [CopG μM]	$S_{20,w}$	% total OD	$S_{20,w}$	% total OD	$S_{20,w}$	% total OD	$S_{20,w}$	% total OD	$S_{20,w}$	% total OD
DNA alone	1.76	2.35	3.58	87.5	5.29	7.44				
8:1 [2.4]			3.55	20.5			5.43	77.8	8.06	2.6
16:1 [4.8]			2.25	4			5.71	91.9	8.05	3.3

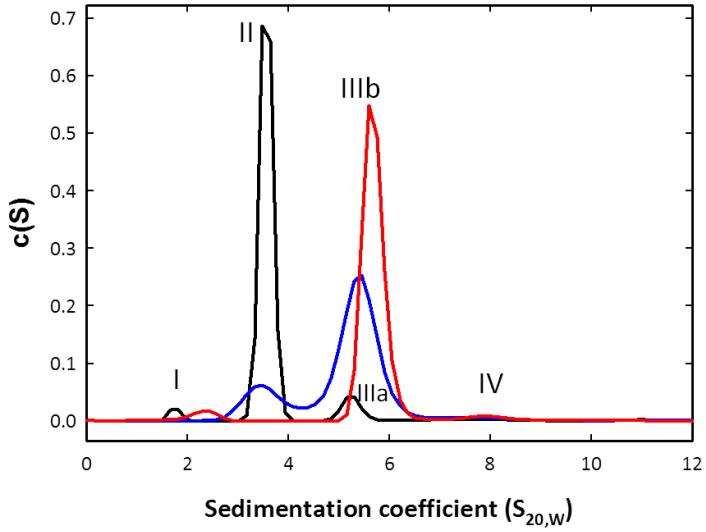


Figure 22. Distribution of the sedimentation coefficients, obtained by the Lamm equation analysis, of CopG combined with its operator DNA. Coefficients were corrected for water at 20 °C. Data of the WT oligonucleotide without (*black line*) and with 2.4 µM (*blue*) and 4.8 µM (*red*) CopG in buffer containing 125 mM NaCl were taken at 48000 rpm and 20 °C.

2. Stoichiometry of CopG binding to its operator DNA

To measure the stoichiometry of CopG bound to its operator, WT oligonucleotide at 0.3 µM of final concentration was combined with CopG at increasing CopG-to-DNA molar ratios and analyzed by sedimentation equilibrium (Figure 23B). Data were collected at 257 nm, a wavelength selected to maximize the contribution of DNA to the absorbance. Therefore, the sedimentation equilibrium gradients obtained corresponded only to DNA or CopG-DNA complexes. See *M&M* (Part II, 7) for details. In parallel with the equilibrium sedimentation assay, an EMSA was performed to analyze the fraction of free and complexed DNA in the different samples (Figure 23A).

The equilibrium sedimentation gradients obtained for individual samples allowed the estimation of the average buoyant molar masses of the unbound and bound DNA pool. This, in turns, allowed us to obtain the number of CopG molecules bound per DNA molecule (*M&M*, Eq. 7), as well as the amount of free protein in the sample (*M&M*, Eq. 9). In Figure 23C, the average number of CopG molecules, represented in dimers ($CopG_2$), bound per DNA molecule is plotted against the concentration of free protein. Under the experimental conditions employed, including the presence of heparin (500 µg/ml), the best-fit to Hill equation (*M&M*, Eq. 10) showed an n value of 3.97 ± 0.15, which is essentially consistent with our working model, where $n_{max} = 4$. That is to say, the maximum number of dimers bound to the wild-type operator DNA is four. Therefore, this analysis demonstrated that the slowest-migrating specific complex observed in EMSA actually corresponds to four CopG dimers bound to the operator DNA.

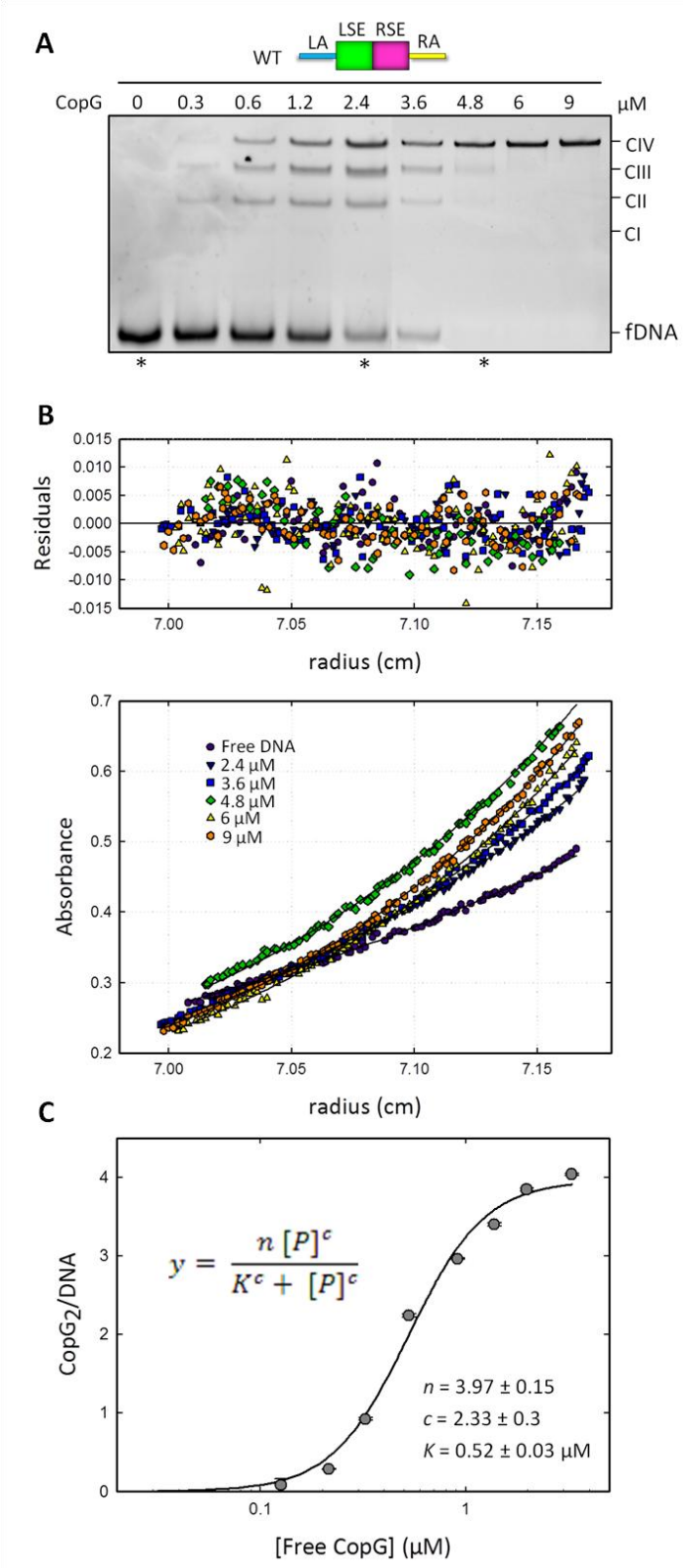


Figure 23. Sedimentation equilibrium of CopG with its operator DNA. (A) EMSA of samples taken from the mixtures prepared for sedimentation equilibrium analysis. WT oligonucleotide (0.3 µM) (*upper colored diagram*) was mixed with increasing concentrations of CopG and allowed to equilibrate for 10 min at room temperature before loading onto the gel. *Asterisks*, samples equivalent to those used in the sedimentation velocity analysis (Figure 22).
(B) Gradients of DNA (0.3 µM) alone (*black circles*) and with several concentrations of CopG equilibrated in buffer containing 115 mM NaCl. Data were taken at 9500 rpm and 20 °C. Continuous lines represent the best-fit single-species model that accounts for the experimental data. The upper panel shows the best-fit residual distribution.
(C) Isotherm of CopG binding to the WT oligonucleotide by sedimentation equilibrium. Error bars represent standard deviations. The *x-axis* corresponds to the free protein concentration [*P*]. The *y-axis* corresponds to the average number of protein dimers bound per DNA molecule. Continuous line represents the best-fit of the experimental data to the equation inside (Eq. 10), where *n* is the total number of equivalent sites suitable for binding; [*P*], the free protein concentration; *c*, the Hill coefficient; and *K,* the concentration of free protein that gives half-maximal binding.

On the other hand, the Hill coefficient (*c*) is a reliable parameter to estimate cooperativity. When $1 < c < n_{max}$ the cooperativity is positive. Therefore, in this case, the value $c = 2.33 \pm 0.3$, strongly suggested a great cooperative effect of CopG for nucleating its entire operator DNA.

These results together represented the first quantitative evidence for the association state of the repressor CopG with its operator DNA. They confirmed that the specific nucleoprotein complex is formed by binding of four protein dimers to the operator of the *copG-repB* operon with a high cooperativity.

3. Relative affinity of CopG for the different operator binding subsites and the role of cooperativity in complex formation

As mentioned in *Introduction*, the operator of CopG has been defined to consist of four binding subsites: LA, LSE, RSE, and RA. Both central elements (LSE and RSE) share features, namely the presence of a 5'-TGCA-3' box in each of them, that could be considered as basic to the protein affinity; however, the binding of CopG to the side subsites, which lack the precise 5'-TGCA-3' sequence, has also a critical role. The stoichiometry study of the interaction of CopG with its target DNA demonstrated that four CopG dimers bind to the operator. This, together with crystallographic data arising from CopG bound to the SE, suggested that one protein dimer was bound per binding subsite of the operator in the saturated specific nucleoprotein complex. Therefore, it was interesting to discern the importance of each operator subsite for the protein affinity. Moreover, this study provided the key to the understanding of the dual transcription repression mechanism mediated by CopG; specifically the active dissociation of the stable RNAP-P_{cr} promoter complexes. See *Discussion*.

A series of oligonucleotides either harboring or lacking a single binding subsite were chemically synthesized (Figure 24). The sequence 5'-GACTTGA-3' (reading from the center of the SE) was employed to substitute the RSE and the LSE in the corresponding mutant operators lacking either subsite (RSE^- and LSE^-). The RA and LA were replaced with the sequence at either side of the *Eco*RI site of the pBluescript SK+ plasmid vector (STRATAGENE), where the SE of the CopG operator has been cloned (unpublished results). In the non-specific (NS) oligonucleotide, the sequence of all four binding subsites was substituted accordingly. The relative affinity of binding of CopG to the different mutant operators was determined by competitive EMSAs (Figure 25) wherein CopG was incubated with a mixture of two DNAs, namely the 239-bp fragment containing the wild-type sequence of the operator, and each of the 55-bp operator variants. See in *M&M* (Part II, 2.2) for details. Results are summarized in Table 19.

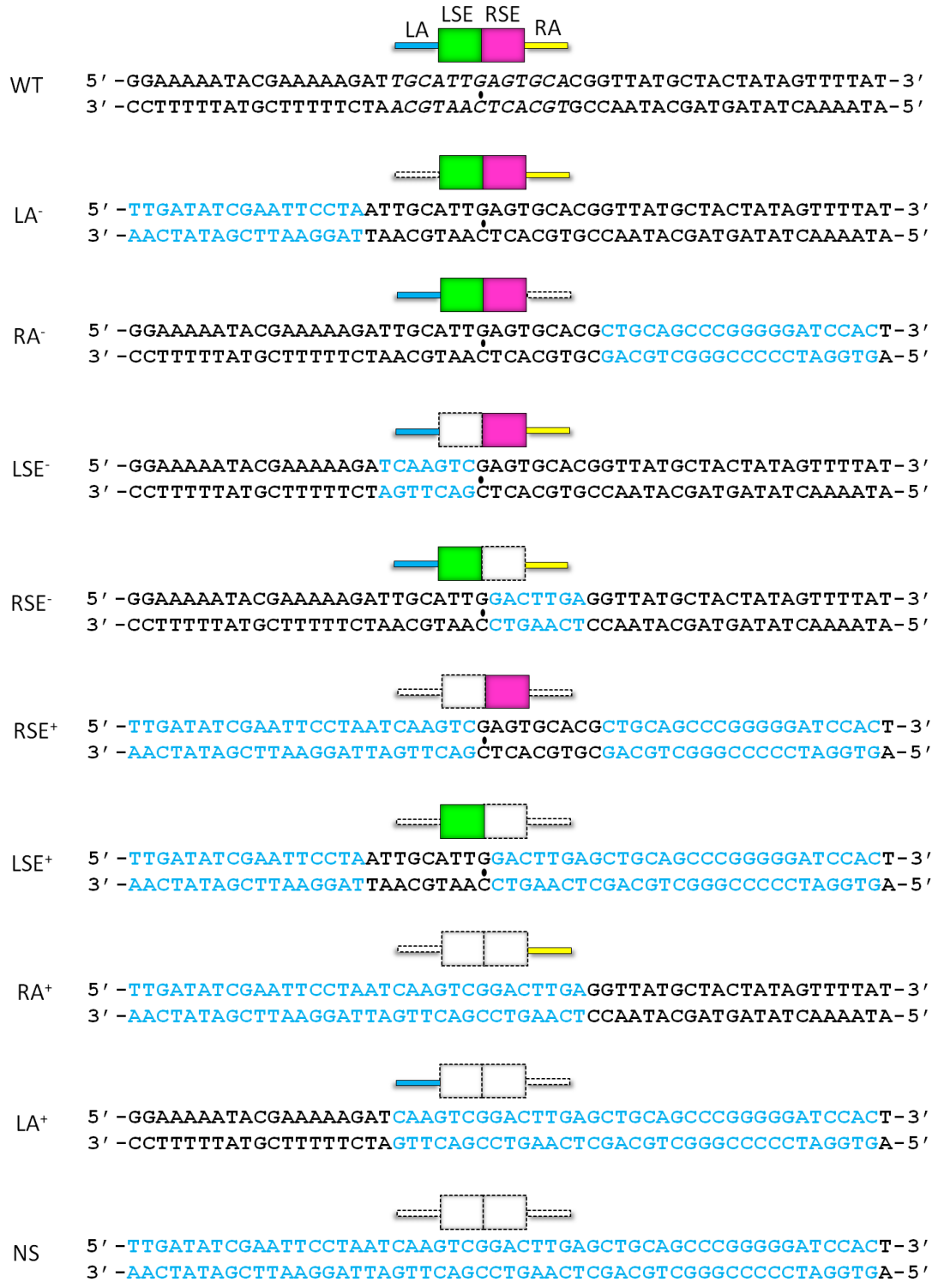
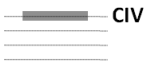
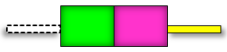
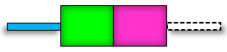
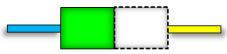
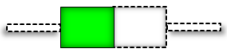


Figure 24. Schemes and nucleotide sequences of the operator variants employed to determine the relative affinity of CopG binding to each subsite. The 55-bp WT operator corresponds to the pMV158 sequence between coordinates 580 to 634. The 13-bp SE is displayed in *italics*, with its two-fold symmetry axis denoted by a *small circle*. Nucleotides modified in the mutant operators are in *blue.* In diagrams, *colored boxes with filled lines* represent the conserved subsites (*blue* for LA, *green* for LSE, *magenta* for RSE, and *yellow* for RA); *white boxes with dashed lines* indicate the absence of the corresponding subsite.

Table 19| CopG binding affinity to different operator variants harboring or lacking only one subsite

	Target DNA	Relative affinity[a]	Complexes[b]
	WT 	1	
Lacking 1 site	LA^- 	0.382 ± 0.046	
	RA^- 	0.150 ± 0.014	
	LSE^- 	0.162 ± 0.022	
	RSE^- 	0.0034 ± 0.0005	
Harboring only 1 site	RSE^+ 	0.0082 ± 0.0014	
	LSE^+ 	< 0.0001	-
	RA^+ 	< 0.0001	-
	LA^+ 	< 0.0001	-
	NS 	< 0.0001	-

[a] The relative affinities calculated for the different operator variants were further normalized with respect to that obtained for the WT oligonucleotide (0.85 ± 0.04). [b] Scheme of complexes at the right shows the highest-order complex that was generated in every case (see Figure 25). In diagrams, *colored boxes with filled lines* represent the conserved subsites (*blue* for LA, *green* for LSE, *magenta* for RSE, and *yellow* for RA); *white boxes with dashed lines* indicate the absence of the corresponding subsite.

From the results obtained with the DNA operators lacking a single binding site, it was evident that substitution of the RSE by the unspecific sequence affected binding of CopG most severely. The affinity of the protein for the RSE^- operator was approximately one-300th the affinity observed for the WT operator. In contrast, a much lower reduction (about one-6th) of binding affinity was obtained when the LSE was replaced by the unspecific sequence. Changing the sequence of either the LA or the RA subsites also caused differential effects on the affinity of CopG binding: whereas replacement of the LA reduced the affinity by about one-3rd, substitution of the RA had the same effect (a reduction in the affinity about one-6th) as removing the LSE (Figure 25, Table 19). These results indicated that the RSE contributed most to the high affinity of binding of CopG to its operator, while the LA had the lowest contribution.

It is important to highlight that the lack of the LSE implies actually the absence of the whole left operator half (LSE and LA subsites), whereas the absence of the RA do not affect other subsite, so that the similar drop in binding affinity of the protein for the operator variants (LSE^- and RA^-) does not reflect necessarily that the LSE and RA subsites contribute equally to overall binding affinity.

To verify whether the RSE was the preferred CopG-binding site, the relative affinity of the protein for modified operators harboring a single subsite was also measured. CopG exhibited detectable specific binding affinity exclusively for the RSE^+ operator, which only harbored the wild-type RSE. The relative affinity of CopG for the other single-subsite operator variants could not be measured, as was also the case for the non-specific (NS) oligonucleotide lacking the four CopG-binding subsites (Figure 25, Table 19). These results show that the affinity of binding of CopG to the LSE^+, LA^+ and RA^+ operator variants is very low and probably approaches non-specific levels. Remarkably, the CopG affinity for the RSE^+ operator was about 2.5-fold greater than the affinity of the protein for the RSE^- operator, therefore showing that the mere presence of the RSE is more important for the binding of CopG than the presence of the other three subsites together. Regarding the complex-formation patterns generated by binding of CopG to each of the operator variants, Table 19 pictures the higher-order complexes observed, which were not necessarily the most-represented complexes, as occurred with the LA^- operator. In this case, as expected, the complex CIII was observed, although for some yet unknown reason, complex CII accumulated. This did not occur with the RA^- operator, whose complexes CII and CIII appeared in the same proportion.

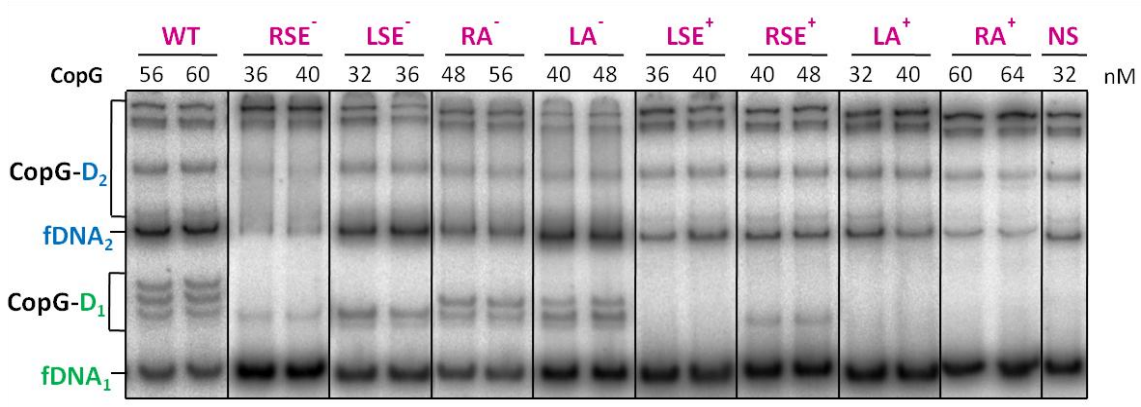


Figure 25. Competitive EMSAs to determine the relative affinity of CopG binding to different operator variants. CopG (at stated concentrations) was incubated at room temperature with a mixture of the indicated operator variant (DNA_1/D_1) and the 239-bp wild-type operator DNA fragment (DNA_2/D_2). Complexes generated by binding of CopG to each DNA ($CopG\text{-}D_1/CopG\text{-}D_2$) as well as unbound DNA ($fDNA_1/fDNA_2$) are indicated.

Summarizing, these results allowed us to conclude that the RSE is the binding site preferred by CopG within the operator, that is to say, the RSE is the primary binding site of the protein. Moreover, the RSE is the only subsite whose presence guarantees specific binding of CopG. Since the RSE is requiered for specific and high-affinity complex formation by CopG, it appears that accesibility of this subsite is an essential factor for the binding of this protein to its operator. In addition, the variable and complex CopG binding patterns observed with the operators lacking any single subsite also demonstrated that all four of the sites are required for proper CopG binding *in vitro*, and that removal of any one subsite perturbs the normal pattern of complex-formation observed with the intact operator. Importantly, CopG binding to subsites other than the RSE (namely LSE, LA or RA) is shown to depend not only on the nucleotide sequence of the site, but also on the inter-dimeric interactions with protein molecules bound to adjacent subsites, hence demonstrating the relevance of cooperativity. The fact that the global affinity for the DNA operator increases as the number of adjacent subsites does, suggests that the nucleoprotein complexes acquires stabilization through protein-protein interactions.

CHAPTER II

Structural features of three CopG variants

1. Two interesting and challenging mutant CopG proteins

Some years ago, we isolated a series of high-copy-number pMV158 derivatives carrying mutations in the *copG* gene. One of them, pLS1*copG7*, has a C to A transversion in position 743 and the other, pLS1*copG8*, possesses a transition G to A in position 728 (coordinates of the pMV158 sequence). Both mutant plasmids encode CopG protein variants unable to repress efficiently the transcription of the *copG-repB* operon (Acebo, Alda *et al.* 1996) and therefore, exhibit increased *N*-values compared to the wild-type plasmid. As an example, plasmid pLS1 carrying the wild-type *copG* gene has a copy number of about 20 in *S. pneumoniae* (del Solar, Acebo *et al.* 1995) versus the *N*-values of about 130 of the plasmids carrying the mutant genes *copG7* and *copG8* (Acebo, Alda *et al.* 1996). The molecular mechanisms underlying the deficiency of these mutant proteins to represses efficiently the transcription of the *copG-repB* operon were unknown until now. Throughout this thesis, the products of mutant genes *copG7* and *copG8* will be named $CopG_{A30E}$ and $CopG_{G25E}$, respectively. $CopG_{A30E}$ possesses a glutamic acid (E) instead of an alanine (A30) at the N-terminus of the α2 helix; and $CopG_{G25E}$ has a glutamic acid (E) to glycine (G25) mutation in the turn between both α-helices. The three-dimensional structure of the wild-type protein (hereafter $CopG_{WT}$) bound to the SE (Gomis-Rüth, Solá *et al.* 1998a) shows that the residues altered in the mutant proteins are located at the dimer-dimer interface (Figure 26). As reported for other RHH proteins, these protein-protein interactions are most likely necessary to stabilize the nucleoprotein complexes, thus rendering both CopG mutants an important tool for the detailed study of the basis of the cooperative binding of the protein.

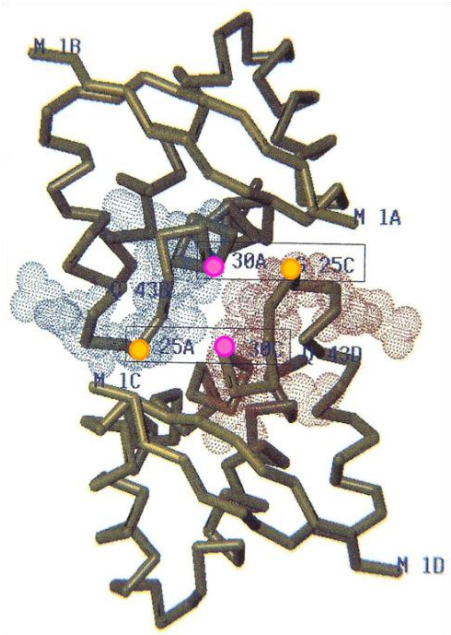


Figure 26. Interaction surface between two CopG dimers bound to the half-sites of the SE. Residues G25 (*yellow circles*) and A30 (*pink circles)* altered in CopG mutants are displayed.

2. Synthesis and purification of the wild-type and mutant CopG proteins

The production and purification of all CopG proteins was performed by a combined TAGZyme-IMAC strategy in order to guarantee the efficient production of highly pure and homogeneous recombinant protein samples suitable for NMR studies. Figure 27 displays a summary of the strategy followed. See *M&M* (Part I, 3) for details.

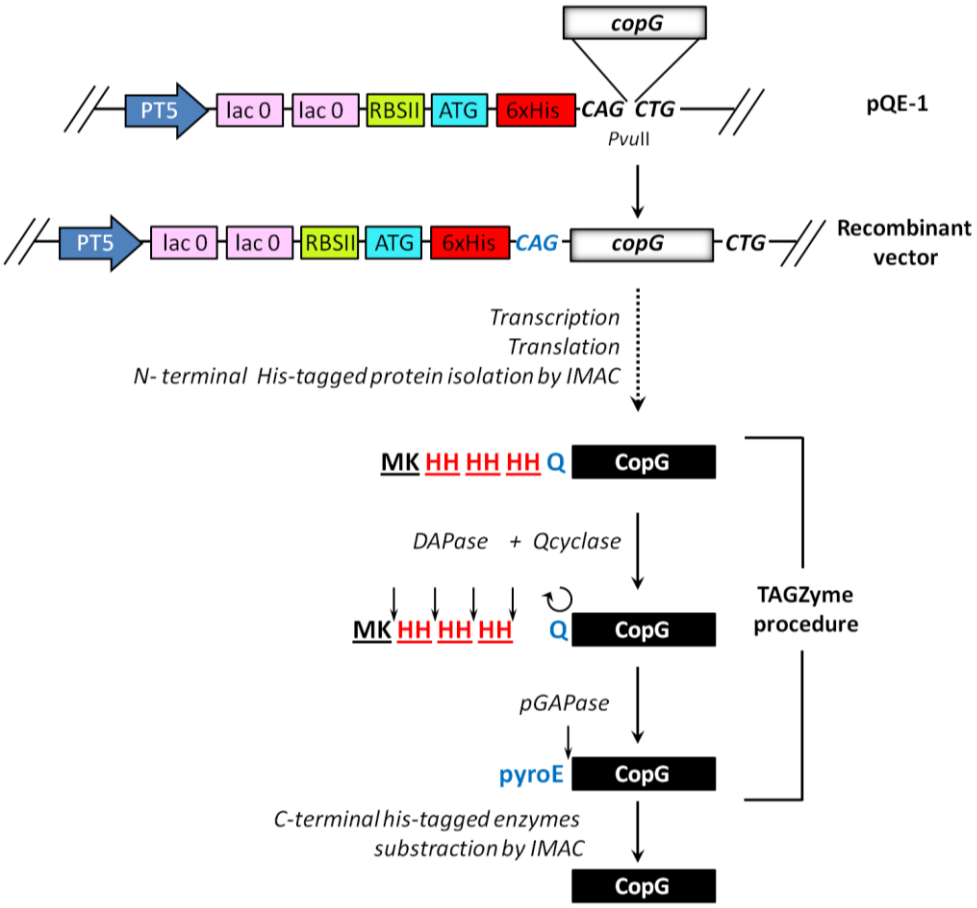


Figure 27. Scheme showing briefly the strategy for the cloning of *copG* genes and purification of CopG proteins. The functional structure of the pQE-1 expression vector and the site of insertion of the *copG* gene are displayed. After cloning, the *Pvu*II site disappeared leaving an upstream triplet encoding for a glutamine (Q). Glutamine is later cyclated (pyroE) and serves as a stop point for the cleavage of dipeptides during the enzymatic procedure of TAGZyme.

Genes *copG*, *copG8* and *copG7*, encoding respectively $CopG_{WT}$, $CopG_{G25E}$ and $CopG_{A30E}$ were separately cloned into the commercial TAGZyme pQE-1 expression vector. Plasmid pQE-1 is a high-level expression vector for production of 6xHis-tagged proteins in *E. coli* (Figure 19) that belongs to the pDS family of plasmids (Bujard, Gentz *et al.* 1987). These low-copy plasmids have an optimized promoter-operator element consisting of phage T5 promoter (recognized by the *E. coli* RNA polymerase) and two *lac* operator sequences, thus ensuring the efficient promoter

repression in the absence of the inductor. In addition, they contain a synthetic ribosomal binding site (RBSII) for high translation rates. Figure 28A shows the recombinant plasmids obtained by cloning of the *copG* gene variants, which were introduced in *E. coli* M15 [pREP4] and selected for ampicillin resistance. The expression of *copG* genes was induced by adding IPTG (Figure 28B), and overall conditions were optimized for a large-scale overproduction of the corresponding proteins.

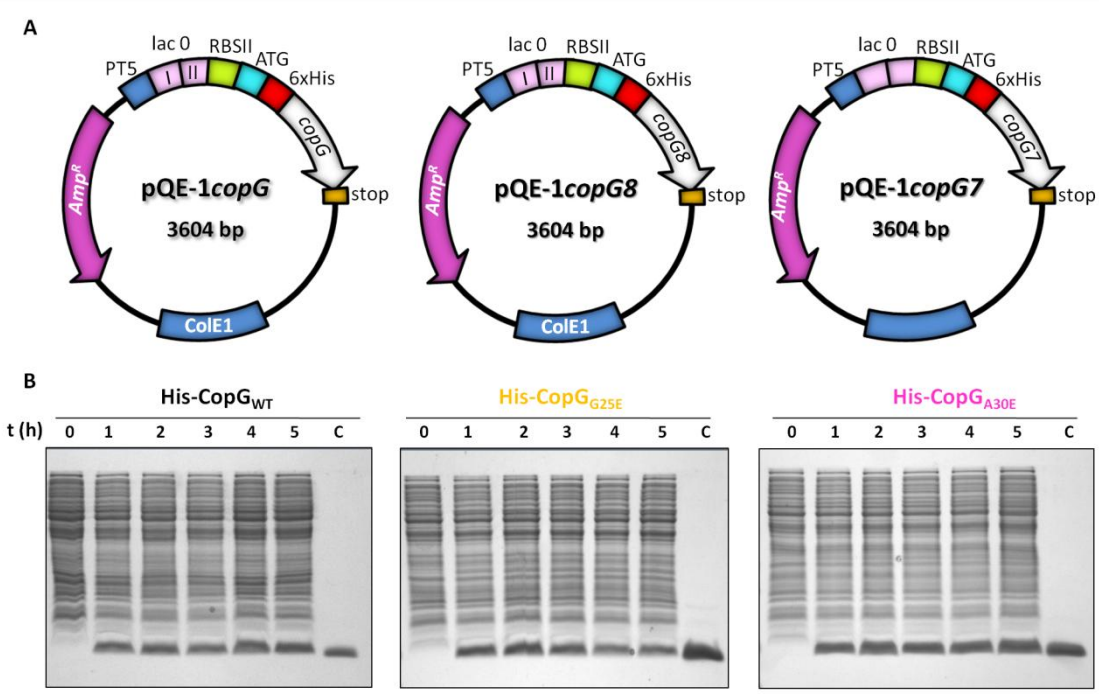


Figure 28. pQE-1 recombinant plasmids and expression of *copG* genes (wild-type and mutants) in *E. coli*. (A) The 138-bp fragments comprising genes *copG*, *copG7* or *copG8* were separately cloned into the single *Pvu*II site of the TAGZyme pQE-1 expression vector. Cloned genes are under control of phage T5 promoter (PT5). Ampicilin resistance is conferred by β-Lactamase gene *bla*. *Lac 0*, *lac* operator; *RBSII*, synthetic ribosome-binding site; *ATG*, start codon; *6xHis*, 6xHis-tag sequence; *ColE1*, origin of replication. (B) Samples were taken from cultures before (lane 0), and 1 to 5 h (lanes 1-5) after induction with 1 mM IPTG. *C*, His-CopG$_{WT}$ purified previously with TAGZyme kit as control.

Proteins synthesized by using the TAGzyme system are tagged with six histidines at the N-terminal end that can be easily removed following an enzymatic treatment. After being purified by IMAC and gel filtration, His-tagged proteins were subjected to the enzymatic process of the TAGZyme system (QIAGEN) for the removal of the N-terminal His-tag. The method is based on the use of dipeptidyl aminopeptidase I (DAPase enzyme) alone or in combination with glutamine cyclotransferase (Qcyclase enzyme) and a pyroglutamyl aminopeptidase (pGAPase enzyme) (Figure 27). In the enzymatic process, N-terminal amino acids are cleaved off as dipeptides by DAPase until a stop point is encountered. Stop points are certain amino acids in defined positions within a dipeptide. The *copG* genes were cloned into the pQE-1 vector in such a way

that a stop point of glutamine (Gln, Q) residue was introduced into the position between the last cleavable dipeptide of histidines and the first authentic amino acid of the target protein (Figure 27). Thus, DAPase cleavage was performed in the presence of the Qcyclase enzyme, which catalyzes the cyclation of the N-terminal Gln-residue to pyroglutamate (pyroE). The pyroE stops the cleavage of dipeptides by DAPase. Finally, the target protein was obtained by cleaving off this N-terminal pyroglutamyl residue through the action of the third enzyme, pGAPase (Figure 27). After completion of the digestion, the enzymes, which carry His-tags at the C-terminus, were also removed by IMAC.

Figure 29 shows the different species of $CopG_{WT}$ that can be isolated at distinct points throughout the procedure (similar species can be isolated for the $CopG_{A30E}$ and $CopG_{G25E}$ mutants). It is worth noting that the new procedure of production and purification of CopG proteins represents an improvement not only in terms of the high throughput and ease of the method but also in purity of the samples obtained as compared with the routinely procedure described (Gomis-Rüth, Solà *et al.* 1998b), where the presence of contaminant proteins could be observed.

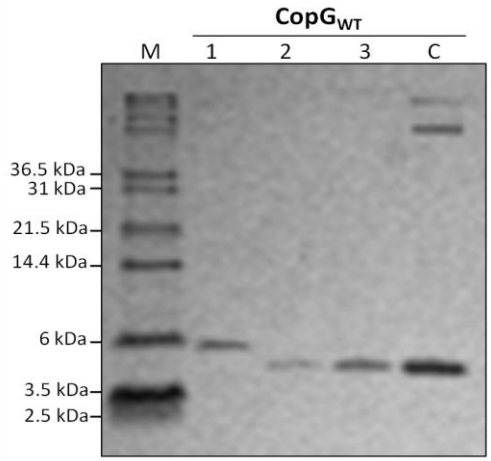


Figure 29. CopG species obtained while performing the TAGZyme protocol. His-$CopG_{WT}$, having the six histidines at the N-terminus (*lane 1*); pyroE-$CopG_{WT}$, bearing the pyroglutamyl residue arising from the cycling of the glutamine stop-point (*lane 2*); and, native $CopG_{WT}$ (*lane 3). C*, $CopG_{WT}$ purified as described by (Gomis-Rüth, Solà *et al.* 1998b) used as control. *M*, molecular weight marker. SDS-PAGE in 16% PAA gels.

3. Secondary structure of the wild-type and mutant CopG proteins

The secondary structure of the mutant proteins was inferred from both Circular Dichroism (CD) spectroscopy and NMR ^{13}C chemical shift analysis. The results from both techniques were consistent and, in particular, the ^{13}C chemical shift analysis demonstrated that the amino acid substitutions did not alter the secondary structure elements of mutant proteins (Figure 30) as compared to $CopG_{WT}$.

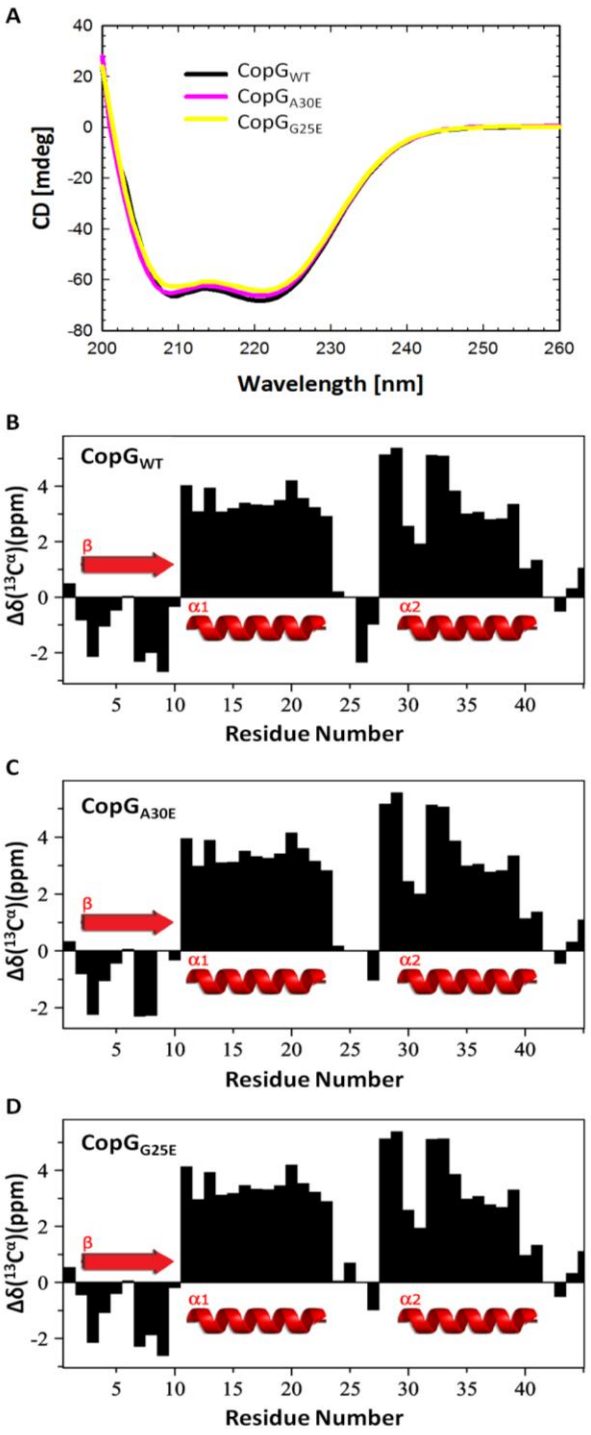


Figure 30. Secondary structure elements of CopG proteins. (A) CD spectra-overlay of protein CopG variants. Far-UV spectra were recorded in 1-mm optical path length quartz cells at a protein concentration of 100 µM in PBS, pH 7.4 and at 20 °C. The difference between the observed $^{13}C^{\alpha}$ chemical shift and the respective random coil value is plotted against the CopG residue-sequence of the wild-type protein (B), $CopG_{A30E}$ (C) and $CopG_{G25E}$ (D).

The CD spectroscopy is a widely used approach to determine the composition of secondary structure elements in proteins by measuring the differences in the absorption of left-handed and right-handed polarized light that result from structural asymmetry. The absence of a regular structure results in zero CD intensity, whereas an ordered structure results in a spectrum that can contain both positive and negative signals. In this case, the secondary structure content of proteins can be determined because of the asymmetry of the peptide bond throughout the

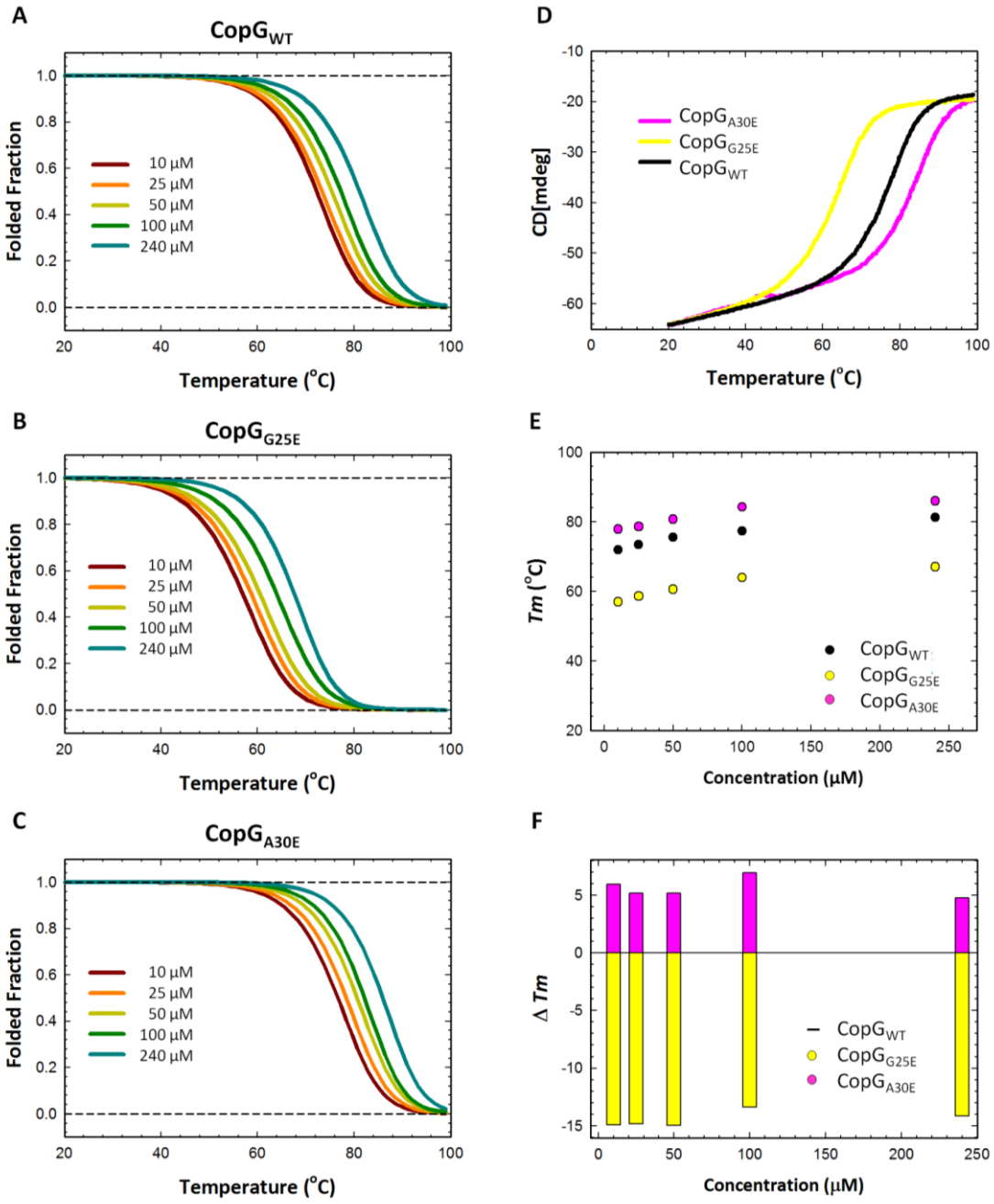
residues. It was observed that, for all the protein concentrations tested (from 10 to 240 µM), the CD spectra in the far-UV (200-260 nm) of both mutant CopG proteins overlapped perfectly the spectrum of the wild-type protein (Figure 30A). On this basis, we inferred that $CopG_{G25E}$ and $CopG_{A30E}$ also consist of one β-strand and a very marked α-helical contribution evidenced by the minima at 209 nm and 222 nm as published for the wild-type protein (Acebo, Alda *et al.* 1996).

On the other hand, NMR chemical shifts statistics are another frequently used tool to predict regions of secondary structure of proteins. In NMR, the chemical shift (δ) for a specific atom arises because electrons in the outer atomic shells and in the chemical bonds generate small magnetic fields that add to or subtract from the external magnetic field B^0. As a result, the local value of the magnetic field depends on the location of the nucleus in the molecule and on the details of the electronic structure. So far, a large number of NMR studies have correlated chemical shifts to secondary structure elements in proteins (Wishart & Sykes 1994). In particular, it has been found that the ^{13}C NMR chemical shift of the C^{α} of all 20 naturally occurring amino acids experiments an upshield shift (with respect to the random coil value) when in a helical configuration and a downfield shift when in a β-strand configuration. Therefore, the Δδ ($^{13}C^{\alpha}$) difference between the $^{13}C^{\alpha}$ chemical shifts of the amino acids in the protein and their corresponding random coil values will be positive for those residues in α-helices and negative for residues in β-strands. Figure 30B-D shows plots of these differences between the experimental chemical shifts of residues of CopG proteins and the corresponding random coil values. These data unambiguously demonstrate that all three proteins have an extended fold (β-strand) from residues 1 to 9 and a helical configuration for residues 11-24 and 28-41. Unlike CD, which is a low-resolution structural technique, NMR chemical shifts yield site-specific structural information, determining the secondary structure at each specific residue within the protein sequence. It is also important to mention that the NMR relaxation analysis of the three protein variants shows that $CopG_{WT}$, $CopG_{G25E}$ and $CopG_{A30E}$ remain as dimers at all the protein concentrations employed (not shown).

4. Measurement of thermal stability of the wild-type and mutant CopG proteins

The effect of temperature on the secondary structure of proteins was monitored by CD changes at 220 nm in the temperature range of 20 to 99 °C. Different protein concentrations were employed in these assays. Results demonstrate that thermal stability of CopG (both wild-type and mutants) increases as a function of protein concentration (Figure 31). This suggests that the unfolding of CopG is not a unimolecular reaction that implies the equilibrium between folded and unfolded dissociated monomers of protein, but a reaction in which the dissociation of the

CopG dimer and the unfolding of the monomers are coupled, like in Arc repressor (Bowie & Sauer 1989). In addition, reversibility of unfolding processes under these experimental conditions was complete in all cases (not shown).



,

Figure 31. Concentration dependence of the thermal stability of the three CopG variants. Curves of thermal-denaturation of CopG$_{WT}$ (A), CopG$_{G25E}$ (B) and CopG$_{A30E}$ (C) were fitted to a folded-unfolded state model as a function of the temperature increment. (D) Curves of thermal-denaturation of wild-type and mutant CopG proteins (100 µM). Measurements were recorded at 220 nm in the temperature range from 20 to 99 °C in PBS, pH 7.4. (E) Plot of the melting temperature (T_m) values at different protein concentrations shown in Table 20. Errors in the T_m values (± 0.5 °C) were determined from duplicate experiments. (F) Variation in T_m-values of mutant CopG proteins relative to the wild-type.

The curves of thermal denaturation observed for the three CopG variants (Figure 31D) also demonstrate that only two states are significantly populated in the unfolding transition; therefore, data of CD profiles were fitted to a two-state unfolding model in order to estimate the thermodynamic parameters of the transitions (Table 20). See *M&M* (Part II, 8) for details. The temperature at which the native (folded) and denatured (unfolded) states are equally populated in equilibrium is called the melting temperature (T_m) and is directly related to conformational stability, only if, as occurred for CopG proteins, the melting is fully reversible.

Plot of T_m-values showed better the differences in thermal-stability among the protein variants (Figure 31D-F). Aside from the concentration-dependence of thermal unfolding for all proteins, thermal stability also varied depending on the CopG variant. Thus, temperature for midpoints of the two-state transitions for $CopG_{G25E}$ were found to be some degrees lower than temperatures observed for the wild-type CopG (Figure 31E and F), which in turn, were found to be a few degrees lower than melting temperatures observed for mutant protein $CopG_{A30E}$ (Figure 31E and F).

Table 20| Melting temperatures and increments of enthalpy of unfolding observed by CD for the two-state transition of CopG proteins

		Protein concentration				
		10 µM	25 µM	50 µM	100 µM	240 µM
$CopG_{WT}$						
	T_m	71.91 ± 0.02	73.45 ± 0.02	75.51 ± 0.02	77.31 ± 0.01	81.21 ± 0.02
	ΔH	-74836 ± 308	-74146 ± 338	-77639 ± 294	-78693 ± 226	-82746 ± 372
$CopG_{G25E}$						
	T_m	56.99 ± 0.03	58.64 ± 0.04	60.57 ± 0.03	63.96 ± 0.03	67.06 ± 0.02
	ΔH	-63206 ± 252	-63335 ± 351	-64671 ± 292	-67368 ± 325	-72912 ± 266
$CopG_{A30E}$						
	T_m	77.86 ± 0.01	78.61 ± 0.02	80.67 ± 0.01	84.23 ± 0.03	85.95 ± 0.02
	ΔH	-80603 ± 275	-83666 ± 432	-83714 ± 284	-87157 ± 604	-90680 ± 393

T_m, melting temperature expressed in °C. ΔH, enthalpy change expressed in cal/mol K

5. Three-dimensional structure of the wild-type and mutant CopG proteins

The solution three-dimensional structure of CopG proteins was solved by NMR (unpublished data). Results demonstrated that the wild-type protein exhibits a symmetrical dimeric configuration, matching with the structure previously described by using crystallography (RMSD of 0.19 Å) (Gomis-Rüth, Solá *et al.* 1998a). The three-dimensional structure of mutant protein $CopG_{A30E}$ resulted very similar to $CopG_{WT}$, showing a RMSD of 0.18 Å (Figure 32). Therefore, as well as the changed residues in the polypeptide chain did not alter the content of secondary

structure elements (Figure 30), they neither affected significantly the global tertiary conformation. However, as a number of NOEs (nuclear Overhausser effects) have not yet been assigned, structure of protein $CopG_{G25E}$ remains to be solved. In general, all data showed that structures of mutant proteins differ only slightly from that of the wild-type protein. These differences could be major when understanding their behavior with the DNA.

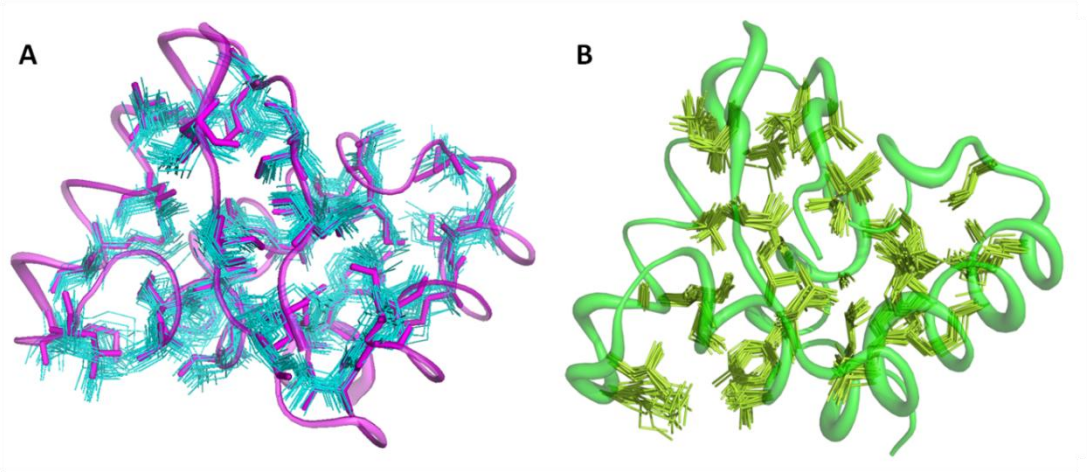


Figure 32. Three-dimensional structure of CopG proteins in solution. (A) $CopG_{WT}$. (B) $CopG_{A30E}$.

Table 21 | Selected NMR data of CopG proteins

	$CopG_{WT}$	**$CopG_{A30E}$**
NMR distance restraints:		
Total NOE	2266	2312
Assigned	2257	2310
Short-range, |i-j|≤1	1570	1561
Medium-range, 1≤|i-j|≤5	348	400
Long-range, |i-j|≥5, intramolecular	125	121
Intermolecular	214	228
Maximal distance restraint violation (Å)	0.14	0.15
Ramachandran plot stratistics:		
Residues in most favored regions (%)	96	98
Residues in additionally allowed regions (%)	4	2
Residues in generously allowed regions (%)	0	0
Residues in disallowed regions (%)	0	0
RMSD from mean coordinates:		
N,Cα,C' of residues 3-41 of both monomers (Å)	0.46 ± 0.08	0.34 ± 0.05
Heavy atoms of residues 3-41 of both monomers (Å)	0.90 ± 0.10	0.72 ± 0.05

NOE, Nuclear Overhauser effect; *RMSD,* Root Mean Standard Deviation

CHAPTER III

Interactions of the three CopG variants with DNA

1. Binding of the His-tagged proteins to the operator DNA

The first evidence of specific binding of the mutant proteins to DNA was obtained by EMSAs using the His-tagged variants and the WT oligonucleotide containing the entire operator. In the case of $CopG_{WT}$, this analysis suggested that histidines affected moderately the cooperative binding (Figure 33A). Even at the highest concentration, His-$CopG_{WT}$ was not able to produce accumulation of the complex CIV, whereas the intermediate-mobility complexes were overrepresented. In contrast, the highly cooperative pattern is observed using $CopG_{WT}$. In this case, the maximal fraction of the intermediate complexes, mainly of the complex CI, was kept low, as though the generation of the higher-order complexes were facilitated once the lower ones were formed. The effect of the histidines attached to the N-terminus of the protein on the cooperativity could be due to its plausible interference with the CopG dimer-dimer interaction surface (Figure 26). Variations between the complex-mobility patterns of both kinds of proteins (with and without the His-tag) are due mainly to the size of the molecule, the slower-migrating complexes corresponding to those carrying the His-tagged protein.

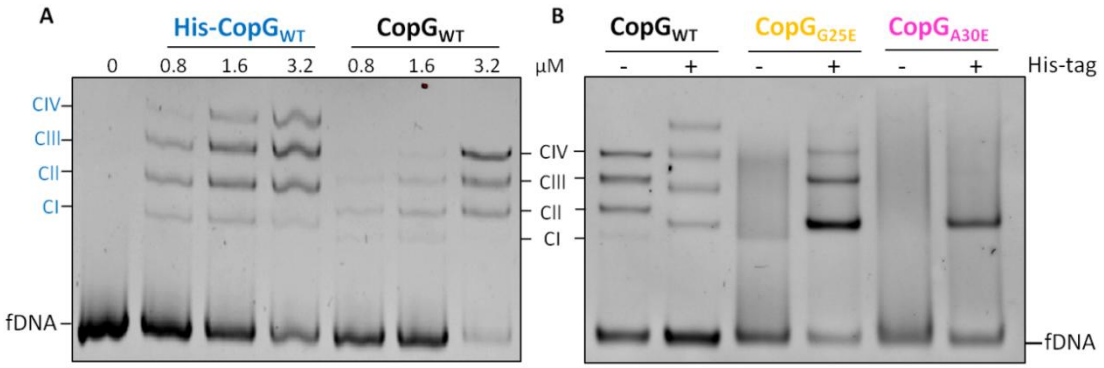


Figure 33. Complex-formation pattern in the binding of the His-tagged proteins to the operator DNA. (A) Complex-migration pattern showing the differences between the complexes generated by $CopG_{WT}$ with and without the His-tag. The indicated concentrations of protein were incubated with the non-radiolabeled WT operator (400 nM). (B) EMSA displaying the effect of the His-tag in the affinity and cooperativity of binding of the three CopG variants to the WT operator (350 nM). In all cases, the protein concentration was 1.4 μM.

In the case of the CopG mutants, the presence of the histidines attached to the N-terminal end seemed to provide somehow a great stabilization of the complex CI, the fastest migrating complex generated by binding to the operator DNA (Figure 33B). Accumulation of the complex CI could be due, at least in part, to the impairment of protein-protein interactions, which might result in turn, from the sum of the negative effects of histidines and of the amino acid modifications of the mutant proteins on the cooperativity. We have observed that in the case of mutant $CopG_{G25E}$ such protein-protein interactions were already severely impaired, while in the

case of $CopG_{A30E}$, high cooperativity of binding to the two subsites of the right half of the operator was kept (*Chapter III*, 4). On the other hand, and based on the fraction of complexed DNA observed at a given protein concentration, it appeared that the presence of the His-tag increased the overall affinity of the mutant proteins for the operator DNA. This increased affinity most likely results from an improvement in the otherwise very low binding affinity of the mutant proteins to the RSE (*Chapter III*, 3), as inferred from the significance of the CI complex generated by either His-tagged mutant protein. In the case of His-$CopG_{G25E}$, formation of complexes with a slower electrophoretic mobility than complex CI was observed, albeit they did not correspond to those generated by the His-$CopG_{WT}$, and arose in an apparent non-cooperative way. These complexes likely result from binding of further molecules to the secondary operator subsites in the absence of cooperative interactions, so that nucleoprotein complexes with conformations distinct from those of His-$CopG_{WT}$, and hence migrating differently, are generated.

2. Analysis of the specific DNA-binding patterns of the CopG variants

The specific complexes generated by binding of the wild-type, $CopG_{G25E}$ and $CopG_{A30E}$ proteins to the entire operator DNA were next disclosed by using greater concentrations of protein relative to that of $CopG_{WT}$. In addition, the binding reactions were performed in the presence of T7-bacteriophage DNA as competitor to favor the detection of specific complexes (compare Figures 33B and 34). Under these experimental conditions, whereas $CopG_{WT}$ gave rise to the four specific complexes of different electrophoretic mobility (CI, CII, CIII and CIV), altered patterns of complexes were observed for the two mutant proteins (Figure 34). Binding of $CopG_{G25E}$ resulted only in the fastest migrating complex (CI), even at high protein concentration. Hence, this mutant seems to be drastically impaired in binding cooperativity. On the other hand, binding of $CopG_{A30E}$ generated a single complex migrating as the complex CII observed for $CopG_{WT}$, which most likely contains a CopG tetramer bound. Since specific $CopG_{A30E}$-operator nucleoprotein complexes of higher order than CII were not detected, some defect in cooperativity was also inferred. As observed, reduced operator binding affinity of both CopG mutants was also suggested from the higher protein concentration required to generate specific complexes compared to the wild-type protein.

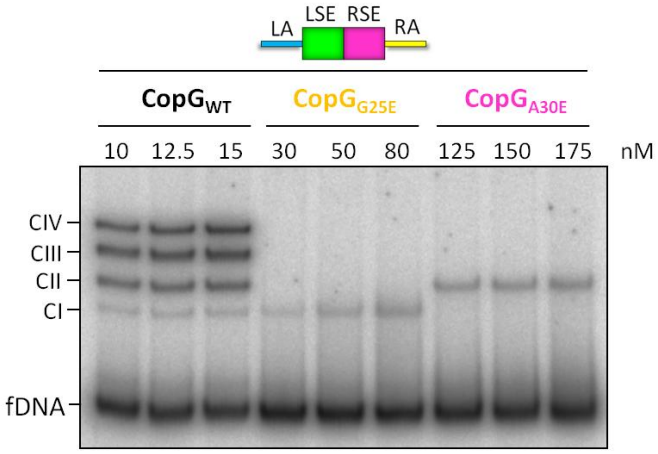


Figure 34. Complex-formation patterns generated by the wild-type and mutant CopG proteins bound to the wild-type operator DNA. Radiolabeled WT operator (2.5 nM) was mixed with proteins (at the stated concentrations) and incubated for 10 min at room temperature in the presence of T7-bacteriophage DNA as competitor. In diagrams, *colored boxes* represent the operator subsites (*blue* for LA, *green* for LSE, *magenta* for RSE, and *yellow* for RA).

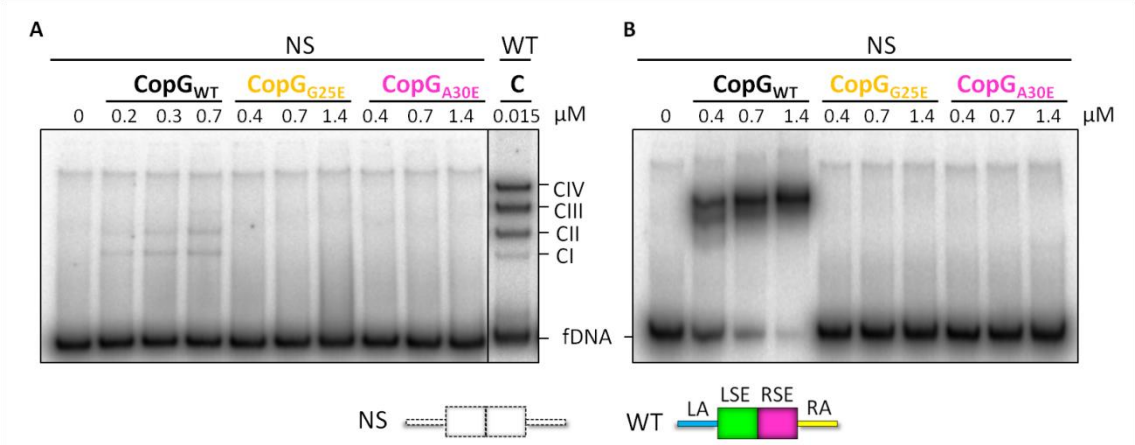


Figure 35. Non-specific binding of the three CopG variants. Radiolabeled NS-oligonucleotide (17 and 34 nM, *A* and *B* respectively) was mixed with the proteins (at the stated concentrations). Binding reactions and electrophoresis were carried out at room temperature in the presence of T7-bacteriophage DNA as competitor (A), or at 0 °C in the presence of 120 µg/ml heparin (B). A sample of CopG$_{WT}$ bound to WT operator at room temperature is shown as control in A (lane C). In diagrams, *colored boxes with filled lines* represent the operator subsites (*blue* for LA, *green* for LSE, *magenta* for RSE, and *yellow* for RA). *White boxes with dashed lines* indicate the absence of the corresponding subsite.

It is worth noting that this reduction in affinity relative to CopG$_{WT}$ was also observed in the binding of the protein mutants to the 55-bp NS oligonucleotide that lacks all four binding subsites of the operator (Figure 35). Whereas the wild-type protein is still able to bind to the non-specific DNA with very low affinity, no complexes were detected with the mutant proteins (Figure 35A). To maximize the affinity for the non-specific DNA, binding reactions were carried

out at 0 °C in presence of heparin (Figure 35B); however, under these experimental conditions, the binding of mutant proteins to DNA neither occurred. These results are consistent with the residues altered in the mutant proteins not being involved in direct interactions with the DNA bases.

3. Binding affinity of the wild-type and mutant CopG proteins to the RSE

Despite their apparent diminished affinity for the operator DNA, it was confirmed by competitive EMSA that the RSE was still the primary site for binding of mutant CopG proteins. A mixture of two DNAs, one of them, a 239-bp DNA fragment harboring the full-length operator, and the other, a 55-bp oligonucleotide either containing or lacking only the RSE (RSE^+ and RSE^-, respectively) was incubated with both mutants. Although very poorly, $CopG_{A30E}$ and $CopG_{G25E}$ only formed complexes with the oligonucleotide conserving the RSE (data not shown), thus demonstrating that the RSE remains the primary site also for mutant proteins. Therefore, it was interesting to measure the binding affinities of the three protein variants for the isolated RSE.

For this purpose, EMSAs using the 18-bp oligonucleotide o(RSE) (Figure 36D), which comprises exclusively the RSE of the operator, were carried out. This oligonucleotide was incubated to equilibrium with increasing concentrations of protein at 0 °C in order to obtain accurately measurable fractions of DNA bound with mutant proteins. See *M&M* (Part II, 4) for details.

From these experiments, the equilibrium dissociation constant (K_D) of the nucleoprotein complex was calculated as the concentration of free protein at which half-maximal binding to DNA occurred. Hence, the lower the K_D the higher the affinity of protein for DNA. Therefore, the affinity of protein for the RSE can be defined as a quantitative parameter related with the inverse of the K_D. Furthermore, since the K_D represents the concentration of free protein, adding a lot more protein than DNA we can assume that as much DNA being bound, the concentration of free protein is almost equal to the total protein concentration added in the assay. It will be reliable provided that the K_D obtained is at least 10 times higher than the concentration of DNA used.

The K_D of the three CopG variants and their representative curves are shown in Table 22 and Figure 36, respectively. $CopG_{WT}$ has a K_D of ~0.8 nM, whereas the mutants showed increases in the dissociation constant value ranging from 10-fold ($CopG_{G25E}$) to 100-fold ($CopG_{A30E}$) and, therefore, proportional decreases in their binding affinity. Oddly, the fraction of DNA bound to $CopG_{A30E}$ was only about 15%. It is worth noting that a 40-fold increase in the dissociation constant value for $CopG_{WT}$ was obtained when these experiments were performed at room

temperature (not shown). At present, we do not know whether the decrease in binding affinity at higher temperatures results from higher dissociation rates or lower association ones. Be it as it may, these results allowed us to demonstrate a diminished affinity of the mutant proteins for the primary binding site, the RSE.

Table 22 | Dissociation constants of the equilibrium between the wild-type or mutant CopG proteins and the primary site

Protein	K_D (0 °C) (nM)	B_{max} (%)	R2	n^a
$CopG_{WT}$	0.76 ± 0.09	82.89 ± 3.86	0.9827	5
$CopG_{G25E}$	8.19 ± 1.75	59.59 ± 2.45	0.9261	4
$CopG_{A30E}$	84.25 ± 12.27	14.99 ± 0.42	0.9698	4

[a] Number of independent experiments used to calculate the K_D.

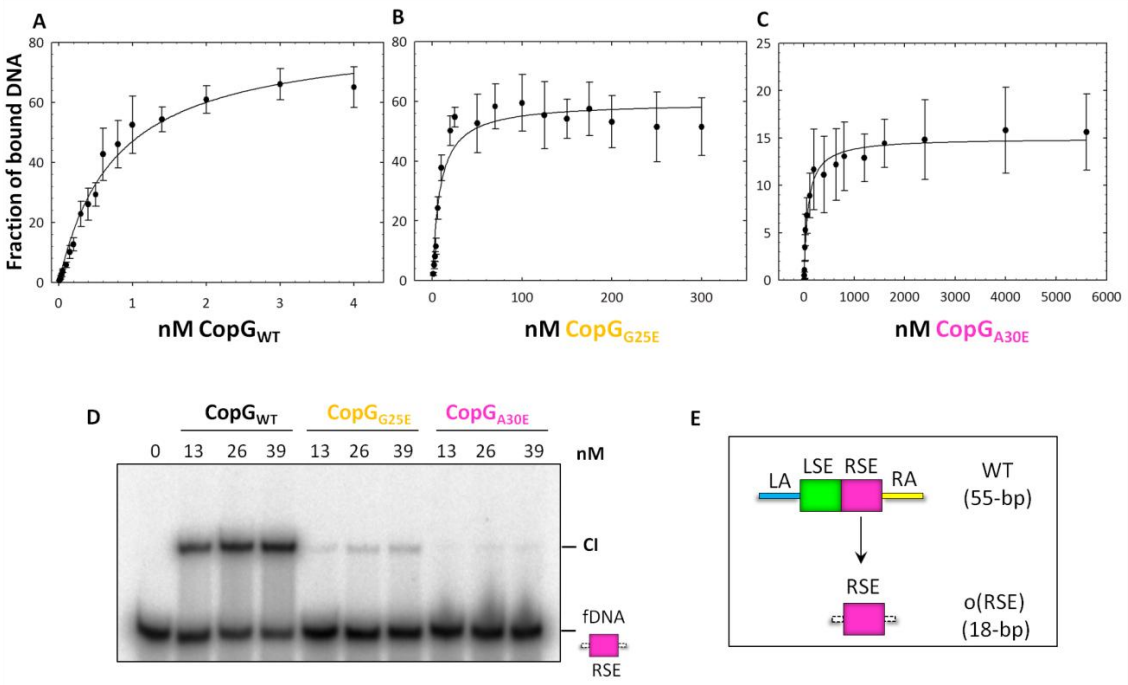


Figure 36. Binding affinities of the wild-type and mutant CopG proteins to the RSE. (A,B,C) Plots of the fraction of DNA bound at different concentrations of $CopG_{WT}$, $CopG_{G25E}$ and $CopG_{A30E}$, respectively. *Dots* represent the average of 4-5 independent experiments, performed by incubating the radioalabeled o(RSE) oligonucleotide (0.05, 0.5 and 4 nM in panels A, B and C, respectively) with increasing concentration of protein, for 30 min at 0 °C. *Error bars* indicate the SD. *Continuous line* represents the best-fit of Eq. 5 to the experimental data. (D) EMSA displaying the differences in binding affinity of the wild-type and mutant CopG proteins to the RSE (3.25 nM). Binding reactions were carried out at room temperature and without additives. (E) Schemes of the full-length operator showing all the subsites and of the 18-bp oligonucleotide comprising exclusively the RSE, which was used in these assays.

4. Binding of CopG proteins to operator variants harboring only two adjacent subsites

It has been demonstrated that the RSE is the unique subsite able to catch the protein by itself (*Chapter I*, Table 19) forming of a nucleoprotein complex stable enough to be detected by EMSA. Therefore, binding of CopG to the other subsites should be mediated by cooperative interactions with a dimer already bound or by interactions between dimers that bind simultaneously to adjacent subsites. On this basis, it was interesting to study the binding of CopG to operator variants harboring exclusively two adjacent subsites, whether or not one of them was the RSE.

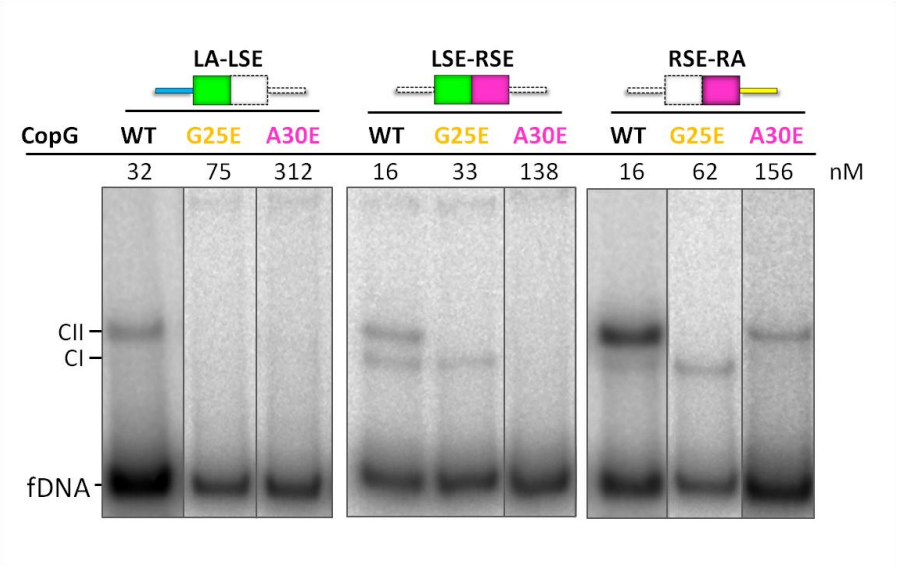


Figure 37. Complex-formation patterns of the wild-type and mutant CopG proteins bound to DNA operators harboring only two adjacent subsites. Radiolabeled 55-bp operators LA-LSE (6 nM), LSE-RSE (3 nM) and RSE-RA (3 nM) were mixed with proteins (at the stated concentrations) and incubated for 10 min at room temperature in the presence of T7-bacteriophage DNA as competitor. Different lanes coming from the same gel were separated by lines if not adjacent. In diagrams, *colored boxes with filled lines* represent the conserved subsites (*blue* for LA, *green* for LSE, *magenta* for RSE, and *yellow* for RA); *white boxes with dashed lines* indicate the absence of the corresponding subsite.

By EMSA, we observed that whereas $CopG_{WT}$ was able to bind to all the two-site operators, both mutant proteins exhibited a selective binding (Figure 37). $CopG_{G25E}$ formed complex CI exclusively with those operators harboring the RSE, regardless of the adjacent element; however, it was unable to generate specific complexes of higher order than CI because, as mentioned previously, this mutant protein has a severe deficiency in cooperativity. On the other hand, the mutant $CopG_{A30E}$ bound exclusively to the DNA that harbored the right operator half (RSE-RA). No complex formation was observed between this mutant protein and the LA-LSE operator, probably because of the absence of the RSE; but no complex was either formed with the LSE-RSE operator, which does conserve the primary site. It is worth noting that despite the

absence of the primary site, complex CII is generated by binding of $CopG_{WT}$ to the operator LA-LSE, most likely as a result of the cooperative interactions established between each other of the dimers bound (Figure 37).

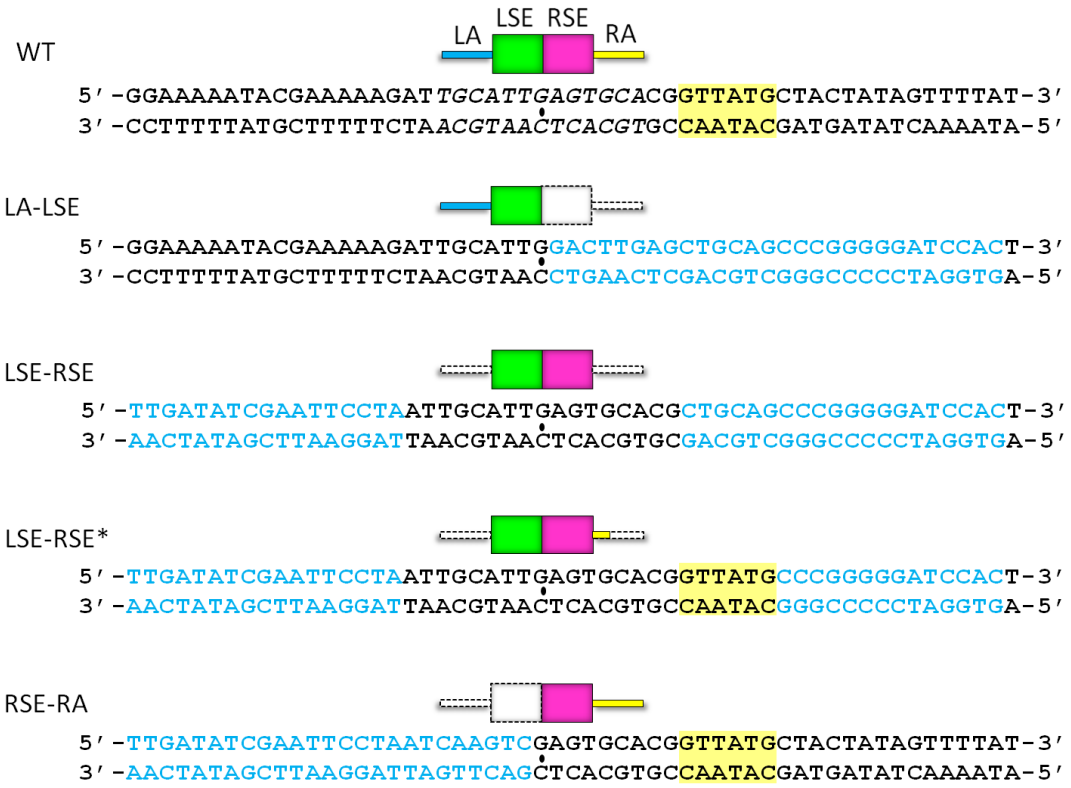


Figure 38. Schemes and nucleotide sequences of the 55-bp operator variants harboring only two adjacent subsites. Sequences of the 55-bp oligonucleotides are shown: WT, carrying the entire operator; LSE-RSE, harboring only the SE; LSE-RSE*, harboring the SE plus the 6-bp next to the RSE; and RSE-RA, comprising only the right operator half. Nucleotides modified in the operator mutants are in *blue*. The 13-bp SE is displayed in *italics*, with its two-fold symmetry axis denoted by a *small circle*. The 6-bp sequence tested for its effect on $CopG_{A30E}$ binding is *yellow-shaded*. *Colored boxes with filled lines* in diagrams represent the conserved subsites (*blue* for LA, *green* for LSE, *magenta* for RSE, and *yellow* for RA); *white boxes with dashed lines* indicate the absence of the corresponding subsite.

The selective binding of $CopG_{A30E}$ to the right half of the operator prompted us to consider whether the sequence 5'-GTTATG-3', adjacent to the right side of the RSE in the wild-type operator, had an essential role in binding of this protein (Figure 38), since this AT-rich sequence is present in the RSE-RA operator as well as in the o(RSE) oligonucleotide containing the RSE alone, but missing in the LSE-RSE one. To test this hypothesis, another operator (LSE-RSE*) was constructed by extending the right boundary of the wild-type sequence in order to include these further bases (Figure 38). The effect of this modification on the binding of mutant $CopG_{A30E}$ was analyzed by EMSA (Figure 39). As it can be observed, $CopG_{A30E}$ remained unable to bind to the

LSE-RSE* operator under the same experimental conditions that prevented binding to the LSE-RSE variant (Figure 39A). In addition, no evident differences were observed in the binding of $CopG_{WT}$ to the LSE-RSE* relative to the LSE-RSE operator (Figure 39B). Hence, the selective binding of $CopG_{A30E}$ might indicate that the interdimeric surface formed by CopG bound to the RSE and RA subsites differs from that described in the co-crystals for the tetramer bound to the SE (LSE and RSE subsites), and that whereas $CopG_{WT}$ is able to accommodate both protein-protein surfaces, $CopG_{A30E}$ can only accommodate that in the right half of the operator.

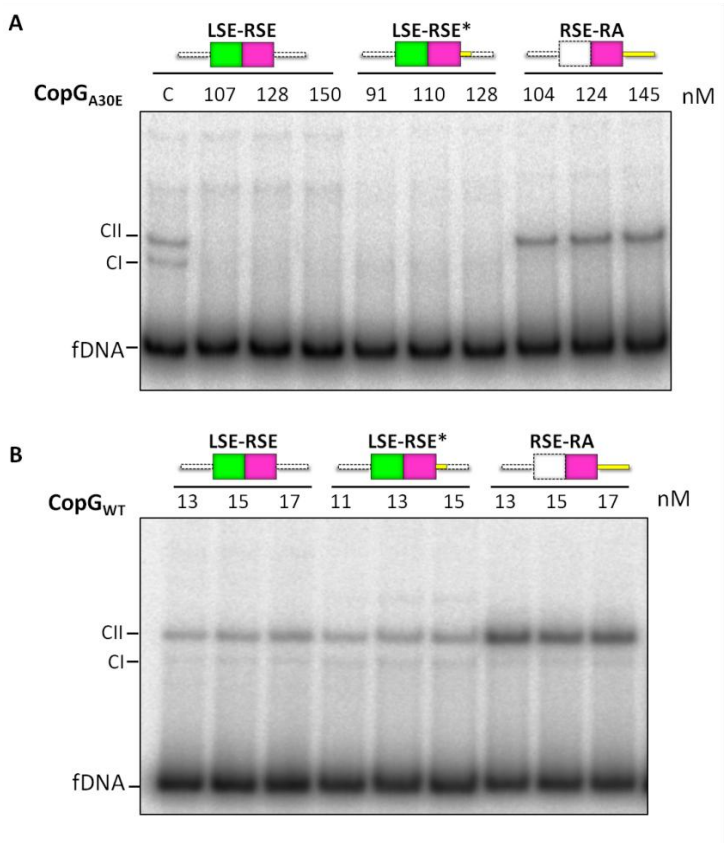


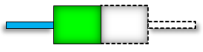
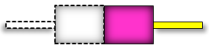
Figure 39. Complex-formation patterns of $CopG_{A30E}$ and $CopG_{WT}$ bound to the LSE-RSE, LSE-RSE* and RSE-RA operators. Radiolabeled 55-bp operators (2 nM) LSE-RSE, LSE-RSE* and RSE-RA were mixed with the indicated concentrations of $CopG_{A30E}$ (A) or $CopG_{WT}$ (B), and incubated for 10 min at room temperature, in the presence of T7-bacteriophage DNA as competitor. Lane *C*, binding of $CopG_{WT}$ (12.5 nM) to the LSE-RSE operator as reference in (A). In diagrams, *colored boxes with filled lines* represent the conserved subsites (*green* for LSE, *magenta* for RSE, and *yellow* for RA); *white boxes with dashed lines* indicate the absence of the corresponding subsite.

5. CopG binds with the highest affinity to the right half of the operator

Since previous results already suggested the existence of a putative interdimeric surface that would be different to that generated by the binding of two dimers to the SE of the operator, we were interested in determining whether CopG binds with higher affinity to the RSE-RA pair than to the pair of central subsites. To address this issue, we measured the binding affinity of CopG

proteins for the whole operator and for operators harboring different combinations of two adjacent operator subsites, relative to the RSE alone. For this, competitive EMSAs were carried out using mixtures of two different DNAs, one of them being a 55-bp oligonucleotide harboring either the entire operator (WT) or only two subsites (LA-LSE, LSE-RSE or RSE-RA), and the other being the 18-bp oligonucleotide containing exclusively the RSE. Each DNA-mixture was incubated separately with the different protein variants and loaded onto PAA native gels (Figure 40). Results are summarized in Table 23.

Table 23 | Binding affinities of CopG proteins to the RSE relative to different operator variants

		DNA operators	**CopG$_{WT}$**	**CopG$_{G25E}$**	**CopG$_{A30E}$**
o(RSE) *vs*	WT		0.0074 ± 0.0013	0.82 ± 0.11	0.018 ± 0.003
	LA-LSE		1.808 ± 0.182	ND	ND
	LSE-RSE		0.297 ± 0.060	1.78 ± 0.38	ND
	RSE-RA		0.041 ± 0.007	0.78 ± 0.16	ND

ND, not determined. In diagrams, *colored boxes with filled lines* represent the conserved subsites (*blue* for LA, *green* for LSE, *magenta* for RSE, and *yellow* for RA); *white boxes with dashed lines* indicate the absence of the corresponding subsite.

The affinity of CopG$_{WT}$ for the RSE alone was about one-130th the affinity observed for the WT operator, as expected from a key role of cooperativity in protein binding. On the other hand, the mutant protein CopG$_{G25E}$ was basically unable to discriminate between the entire operator and the separate RSE, showing a similar affinity for both target DNAs. This result confirmed again that the cooperative properties of CopG$_{G25E}$ are severely impaired, and hence the presence of adjacent subsites has not a contribution to its overall affinity for DNA. In the case of CopG$_{A30E}$, as demonstrated above, this mutant exhibits a quite small affinity for the primary site (Table 22, Figure 36); therefore, under the experimental conditions employed, the complex CI formed by the protein bound to the oligonucleotide containing exclusively the RSE was hardly noticeable. We observed that this mutant protein retains some cooperativity, since its binding affinity for the RSE alone was about one-50th the affinity for the entire operator, albeit CopG$_{A30E}$ binds to only two subsites of it (Figures 34 and 37).

In the case of the operator variants carrying different combinations of two adjacent subsites, the results showed that the binding affinity of $CopG_{WT}$ to the right half of the operator was greater than to the central region, and much greater than to the left operator half (Table 23). Thus, $CopG_{WT}$ showed the highest binding affinity to the RSE-RA operator (25-fold higher than to the RSE), whereas the binding affinity to the LSE-RSE operator was only 3.5-fold higher than to the primary site. This data strongly suggested that the right half of the operator is considerably more significant for binding of the protein than the symmetric central subsites. On the other hand, the affinity for the RSE was 2-fold higher than for the LA-LSE operator; i.e. is the only case in which the affinity for an operator with a single subsite (in this case, the primary) is greater than for a two-site operator (the LA-LSE). Since the RSE is missing in this operator and the protein is unable to bind to a single subsite other than the primary site, this result suggests that two dimers have to bind simultaneously to the left operator half, which is consistent with the observation of only complex CII (Figures 40 and 37). As a consequence, binding of the protein to the LA-LSE operator seems to be highly cooperative (see below). It is also important to highlight that this nucleoprotein complex ($CopG_{WT}$-LA-LSE operator) seemed to migrate slightly faster than those formed on the rest of the two-site operators (Figure 40).

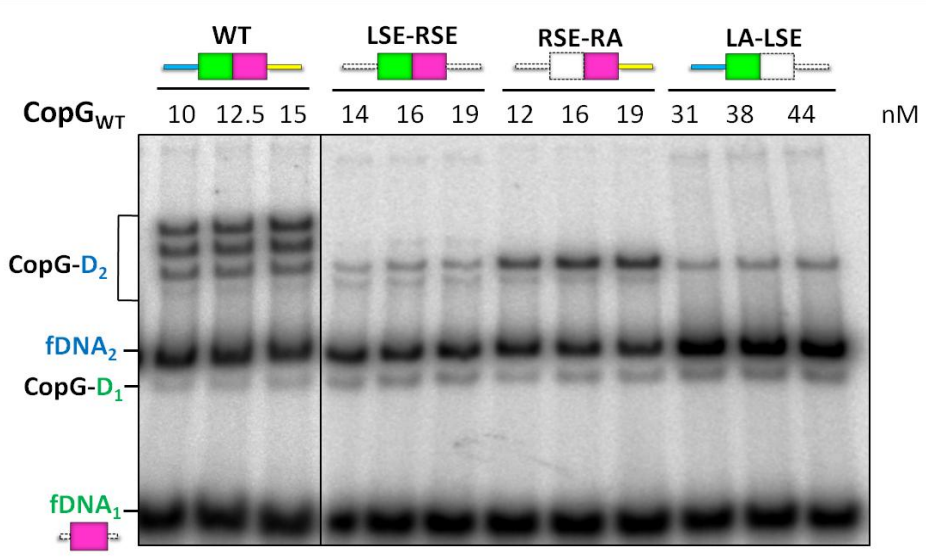


Figure 40. Competitive EMSAs to determine the relative binding affinity of $CopG_{WT}$ to the RSE. $CopG_{WT}$ (at stated concentrations) was incubated at room temperature with a mixture of the indicated operator variant (DNA_2/D_2) and the o(RSE) oligonucleotide (DNA_1/D_1). Complexes generated by binding of CopG to each DNA (*CopG-D_1* or *CopG-D_2*) as well as unbound DNA (*fDNA_1* or *fDNA_2*) are indicated. Lanes coming from different gels run in parallel are separated by a line. In diagrams, *colored boxes with filled lines* represent the conserved subsites (*green* for LSE, *magenta* for RSE, and *yellow* for RA); *white boxes with dashed lines* indicate the absence of the corresponding subsite.

On the other hand, and as expected, binding of $CopG_{G25E}$ to these two-site operators was only observed when using target DNAs that conserved the RSE. The results indicate that $CopG_{G25E}$ does not discriminate among the entire operator, the right operator half and the RSE alone. However, the binding affinity to the primary site was almost 2-fold higher than the binding affinity to the LSE-RSE operator. This result may suggest that for some reason, $CopG_{G25E}$ accommodated worst in the LSE-RSE than in the RSE-RA or in the separate RSE, in spite of all three operator variants harboring the primary site. It is still unknown whether the six nucleotides (Figure 38) adjacent to the RSE, which are present in the RSE-RA operator as well as in the o(RSE) oligonucleotide but absent in the LSE-RSE target DNA, exerts a sort of effect on binding of the mutant $CopG_{G25E}$ to DNA.

Although the binding affinity of $CopG_{A30E}$ to the RSE-RA operator (the only one that binds significantly to this mutant) relative to the RSE was not determined, we hypothesize that it is likely to be as that observed for the entire operator (Table 23), since the left half of the operator is dispensable for $CopG_{A30E}$ binding.

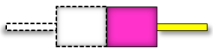
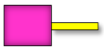
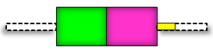
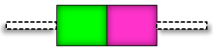
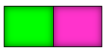
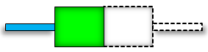
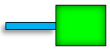
6. CopG binds cooperatively to operator variants harboring only two adjacent subsites

Since the essential role of cooperativity in CopG binding seemed to be clear, we wanted to analyze whether the cooperative interactions between adjacent dimers bound to DNA differed somehow depending on the subsites occupied, and thus affected the affinity of the protein for the operators having different combinations of two adjacent subsites. To get some information, we carried out EMSAs using the operators having only two binding subsites (Figures 41 and 42). Radiolabeled DNAs were mixed with increasing amounts of protein and electrophoresed after equilibrium at room temperature. Binding reactions were performed in the presence of competitor DNA to favor the detection of specific complexes. Bands corresponding to unbound DNA and to the different protein-DNA complexes were quantified in order to graph their relative abundance as a function of the protein concentration (Figures 43 and 44), and to obtain the macroscopic equilibrium constants by applying the model described by (Senear & Brenowitz 1991). See *M&M* (Part II,3) for details.

Comparisons of the macroscopic constants obtained from EMSAs are summarized in Table 24. The results allowed us to confirm qualitatively that cooperativity of CopG binding is positive with all the DNAs employed (Senear & Brenowitz 1991). Moreover, plots of the fraction corresponding to complex CII against the protein concentration described a sigmoid curve,

which is a typical feature of positive cooperativity (Figures 43 and 44). The comparison of the macroscopic constants does not give a quantitative factor of cooperativity; however, this analysis, along with the complex-formation patterns observed, allowed us to infer roughly that cooperativity varies depending on the subsites available in the operator.

Table 24 | Comparison of macroscopic constants obtained from EMSAs with different operator variants

DNA operator			
Size	**Name**	**Picture**[a]	$K_2 / K_1^2/4$[b]
55-bp	RSE-RA		*6714.42 ± 2244.50*
25-bp	o(RSE-RA)		*8458.76 ± 713.14*
55-bp	LSE-RSE*		*3818.41 ± 1063.78*
55-bp	LSE-RSE		*1068.59 ± 197.72*
21-bp	o(LSE-RSE)		*570.35 ± 94.34*
55-bp	LA-LSE		*540274.77 ± 93744.50*
25-bp	o (LA-LSE)		*173268.05 ± 41334.40*

[a] *Colored boxes with filled lines* represent the conserved operator subsites (*blue* for LA, *green* for LSE, *magenta* for RSE, and *yellow* for RA); *white boxes with dashed lines* indicate the absence of the corresponding subsite.

[b] In a two site system, the binding is cooperative if $K_2 > K_1^2/4$. Macroscopic constants K_1 and K_2 were obtained from EMSA, considering only those lanes where complexes CI and CII were clearly separated (Senear & Brenowitz 1991).

As expected, results from EMSAs using the 55-bp operators variants having only two adjacent subsites (LA-LSE, LSE-RSE, LSE-RSE* and RSE-RA) showed that, substitution of the operator subsites by unspecific DNA sequences prevented the formation of those complexes observed by binding of $CopG_{WT}$ to the wild-type operator (*Chapter I*, Figure 21). These observations correlated with the fact that CopG binding is a DNA sequence-specific event. Thus, using these operator variants, the highest-order specific complex mainly observed was the complex CII (Figure 41).

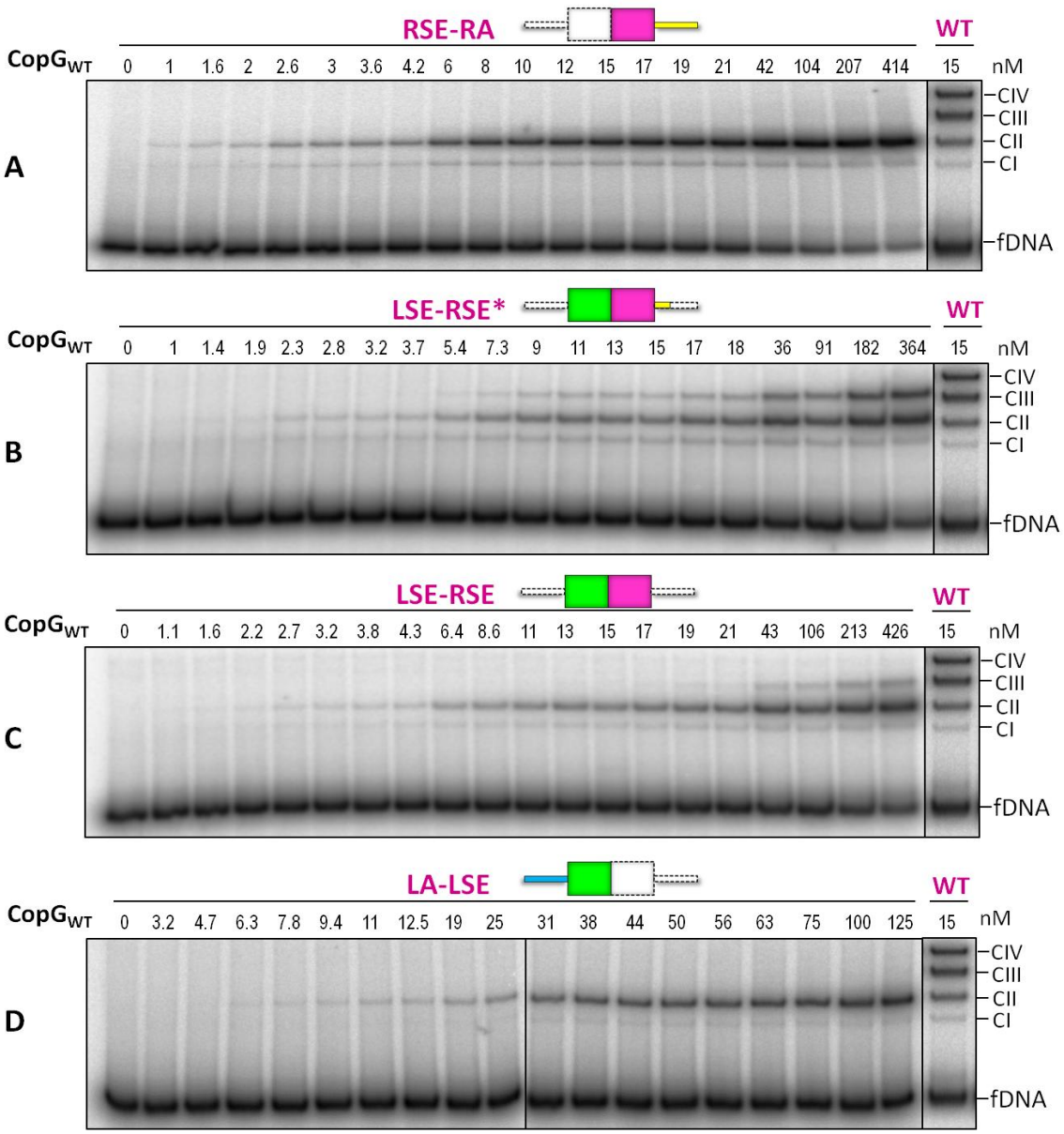


Figure 41. Cooperativity of CopG$_{WT}$ binding to the 55-bp operator variants harboring only two adjacent subsites. Increasing amounts of protein were separately incubated at room temperature with radiolabeled operators (RA-RSE, LSE-RSE* and LSE-RSE at 2 nM; LA-LSE at 6 nM) in the presence of T7-bacteriophage DNA as competitor. In diagrams, *colored boxes with filled lines* represent the conserved subsites (*blue* for LA, *green* for LSE, *magenta* for RSE, and *yellow* for RA); *white boxes with dashed lines* indicate the absence of the corresponding subsite.

In EMSAs, the cooperativity of CopG binding appeared to be higher for the RSE-RA operator than for the LSE-RSE variant (Figure 41, Table 24). Detection of the complex CII formed by binding of protein to the RSE-RA operator even at the lower protein concentrations correlated with the higher affinity exhibited by CopG$_{WT}$ for the right half of the operator (Figure 41A, Table 23). This contrasted with the LSE-RSE operator, where the complex CII was significantly populated only at higher amounts of protein (Figure 41C). Moreover, the maximal fraction of complex CI observed for the RSE-RA operator (0.7%) was lower than those obtained with the LSE-RSE and LSE-RSE*

operators (Figures 41 and 43). It was also observed that cooperativity of $CopG_{WT}$ binding to the LSE-RSE* operator appeared to be higher than that of binding to the LSE-RSE DNA (Figure 41B and C). On the other hand, cooperativity of $CopG_{WT}$ binding to the left operator half (LA-LSE) seemed to be the highest among the two-site operator variants employed, whereas the affinity was the lowest observed. (Figure 41D, Table 24). The complex CI was scarcely detected at any protein concentration (Figure 41D), with a maximum fraction lower than 0.1% (Figure 43D). Due to the absence of the primary site, the binding of two dimers to the two subsites of the left operator half should occur simultaneously.

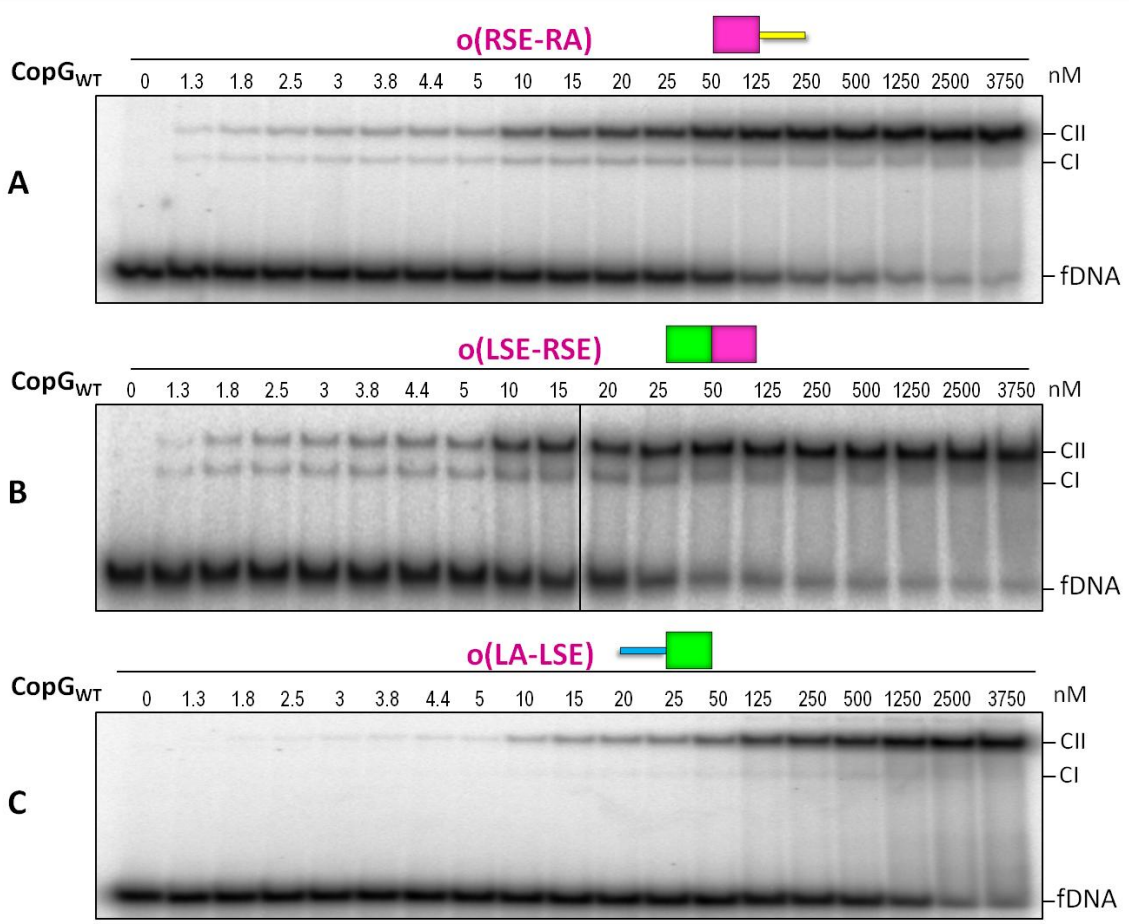


Figure 42. Cooperativity of $CopG_{WT}$ binding to different operator variants consisting only of two subsites. Increasing amounts of protein were separately incubated at room temperature with the radiolabeled operators o(RSE-RA), o(LSE-RSE) and o(LA-LSE) at 2.5 nM in the presence of T7-bacteriophage DNA as competitor. In diagrams, *colored boxes with filled lines* represent the conserved subsites (*blue* for LA, *green* for LSE, *magenta* for RSE, and *yellow* for RA); *white boxes with dashed lines* indicate the absence of the corresponding subsite.

It is worth noting that, in EMSAs using the LSE-RSE and LSE-RSE* operators, a complex that migrated like the complex CIII generated by binding of the protein to the WT operator appeared at high protein concentrations. However, no similar complexes were observed for the RSE-RA or

LA-LSE operators (Figure 41). In the case of the LSE-RSE operator, this complex is likely to be product of the binding of a third dimer to the region substituting the RA. In this mutated half of the operator, a sequence 5'-TGCA-3' very close to the RSE could be promoting the binding of the third dimer, albeit without cooperativity (Figure 38). On the other hand, in the LSE-RSE* operator, the sequence at the joint between the wild-type and mutant nucleotides is a 5'-TGC-3'. This 3-nt region could be part of the recognition sequence in the wild-type RA, hence facilitating somehow the binding of the third dimer to the right half of the operator (Figure 38). In addition, due to its position relative to the RSE, the binding of the third dimer seems to be favored in the LSE-RSE* as compared to the LSE-RSE variant (Figure 41B and C).

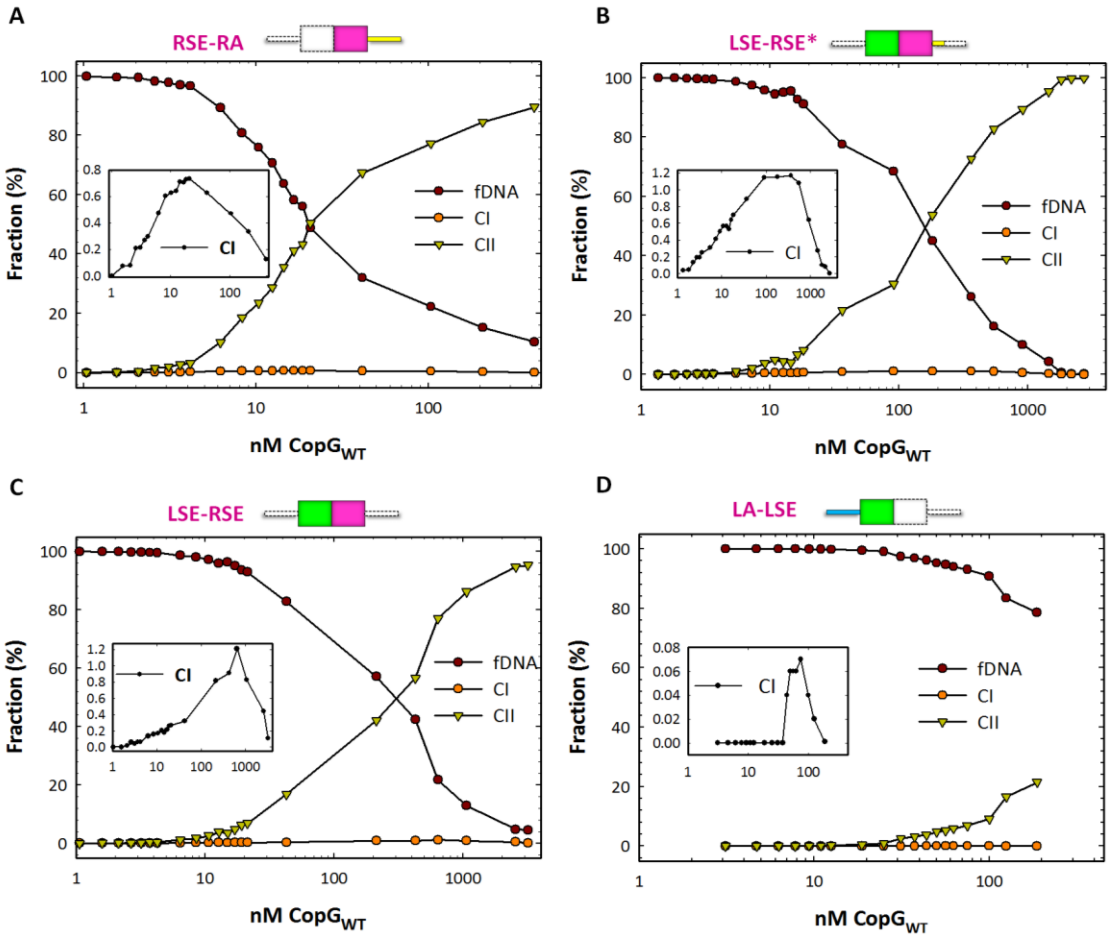


Figure 43. Graphic representation of the cooperative binding of CopG$_{WT}$ to the 55-bp operator variants harboring only two adjacent subsites. Fractions of the different species of DNA were plotted against the protein concentration. Whenever a CIII-like complex was observed, values for complex CII also included the fraction corresponding to the higher-order complex. *Insets*, close-up of fraction of complex CI. In diagrams, *colored boxes with filled lines* represent the conserved subsites (*blue* for LA, *green* for LSE, *magenta* for RSE, and *yellow* for RA); *white boxes with dashed lines* indicate the absence of the corresponding subsite. In *D*, the protein concentration was not high enough to reach a high percentage of DNA bound.

The results obtained with the 55-bp operators variants containing only two adjacent subsites (Figures 41 and 43) were supported by the data coming from the EMSAs carried out by using the short DNA operators consisting only of the same two subsites (Figures 42 and 44). The mere detection of a higher maximal fraction of complex CI with o(LSE-RSE) than with o(RSE-RA) indicates that cooperativity of protein binding to the right half of the operator remains higher than to the SE in the short two-site operators (Figure 44A and B). In the case of the o(LA-LSE) operator, formation of complex CII exhibited the highest apparent cooperativity, as it occurred with the equivalent 55-bp LA-LSE operator variant; a very low maximal fraction of CI (Figure 44C) indicates that the binding of a second dimer is likely to occur almost simultaneously with the first one (Figure 42C).

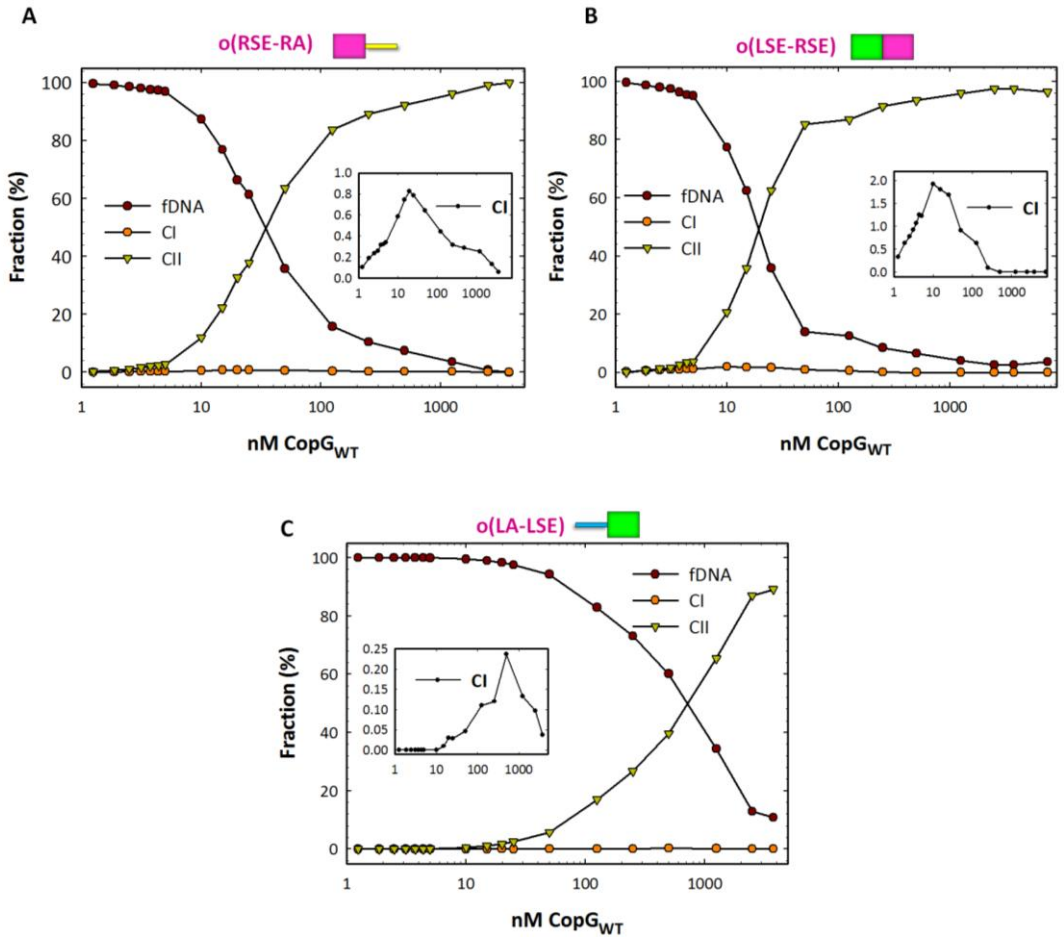


Figure 44. Graphic representation of the cooperative binding of CopG$_{WT}$ to different DNA operators consisting only of two subsites. Fractions of DNA were plotted against the protein concentration. *Insets*, close-up of fraction of complex CI. In diagrams, *colored boxes with filled lines* represent the subsites (*blue* for LA, *green* for LSE, *magenta* for RSE, and *yellow* for RA).

7. Characterization of the nucleoprotein complexes between the wild-type or mutant CopG proteins and the WT-operator DNA

From the results exposed so far, we know that CopG, in the presence of its target DNA, promotes the formation of four complexes, CI to CIV, with different electrophoretic mobilities. Since CopG is a dimer in solution, the fastest-migrating complex might correspond to one protein dimer bound to the RSE in the operator; the subsequent cooperative binding of further dimers to the rest of the subsites (LA, LSE and RA) would generate the higher-order complexes, until reaching the slowest-migrating one (CIV). This pattern of complex formation was prevented by changing the nucleotide sequence of the subsites (Figures 25 and 37) as well as by binding of the mutant proteins (Figure 34). We have shown that $CopG_{A30E}$ only binds to the right operator half, forming a CII-like complex supposed to contain two dimers of protein bound, whereas $CopG_{G25E}$ only generates complex CI, which would correspond to one dimer bound to the primary site, RSE. In regard to this, it was interesting to analyze the contacts established by the three protein variants with the operator DNA. Characterization of the specific complexes generated by the wild-type and mutant CopG proteins bound to the WT oligonucleotide comprising the entire operator was performed by exposition to OH• cleavage and DMS methylation. It is important to mention that the complex CI formed by $CopG_{WT}$ could not be isolated, and therefore characterized, because of the high cooperativity of the protein to generate the next-order complex. Nevertheless, the specific complexes CII, CIII and CIV generated by the wild-type protein, as well as the complex CI generated by $CopG_{G25E}$ and the complex CII formed by $CopG_{A30E}$, were analyzed. Binding reactions were performed at 0 °C to minimize disassembly of the complexes generated by the mutant proteins.

Multiple independent experiments were analyzed to verify the zones that were unclear in the sequencing gels displayed in Figures 45 and 46, which exhibit, respectively, the footprints obtained by protection against OH• cleavage and DMS methylation with the three protein variants. In Figure 47, the results of both techniques are projected on a B-DNA model of the WT operator. For describing the footprints, the top-strand (non-template) sequence is numbered according to the coordinates of the pMV158 sequence, and their complementary bases in the bottom strand (template of the *copG-repB* operon transcription) are denoted with a prime.

7.1. Analysis of the contacts of the CopG proteins with the DNA backbone

The contacts established by the three protein variants with the DNA backbone were analyzed by the protection against OH• cleavage. This technique is based on the principle that if a protein is

bound to DNA, the contacted DNA will be protected from the attack of the OH•, which cleaves DNA by abstracting a hydrogen atom from the deoxyribose sugars along the DNA backbone. Since OH• is exceedingly short lived and reactive, and attacks sites on the surface of the DNA molecule, there is almost no sequence or base dependence in the cleavage reaction. Experimentally, both strands of the WT oligonucleotide were separately radiolabeled, annealed with the non-radiolabeled complementary strand and incubated with proteins. After reaching the binding equilibrium, samples of protein-DNA were exposed to the action of OH•, as indicated in *M&M* Part II, 5.1.1. The treated samples were loaded onto PAA native gels for separation and isolation of the different complexes. The footprints generated were resolved in sequencing gels. The regions where the protein was contacting with DNA appeared as bands with lower intensity than the protein-free regions where the OH• cleaved.

Results showed that the area protected by $CopG_{WT}$ in the complex CIV spanned between the coordinates 584 and 627 (Figure 47). Previous data of OH• footprinting of the CopG-DNA specific complexes generated a 37° C revealed five protected regions on each strand (Hernández-Arriaga, Rubio-Lepe *et al.* 2009). In the present work, the footprinting experiments were performed at 0° C and showed that in the complex CIV, CopG protected five regions in the top strand, whereas only four were observed in the bottom one (Figure 47A). This difference in the protection patterns may result from the effect of temperature on the DNA topology, since at low temperature the twist of the double helix increases (less than 10.5 bp/helical turn). The protected regions (footprints) were named by letters beginning at the 5'-end of the top strand, and served as reference to describe footprints in other complexes. Thereby, protected regions in the top strand were: footprint *a* (from 585 to 589), footprint *b* (from 595 to 600), footprint *c* (from 607 to 611), footprint *d* (from 615 to 619) and footprint *e* (from 624 to 627). The protected zones in the bottom strand were: footprint *a'*+*b*' (from 584' to 595'), footprint *c'* (from 599' to 604'), footprint *d'* (from 611' to 616') and footprint *e'* (from 623' to 625'). The unique hyper-exposition to the OH• cleavage was observed at position 601, and it remained hyper-exposed in the rest of complexes. It is worth nothing that the footprints *a*' and *b*' were not clearly separated; therefore, the whole protected region in the far 3'-end of the bottom strand was named *a'*+*b'* for keeping the stated nomenclature in the published data (Hernández-Arriaga, Rubio-Lepe *et al.* 2009). The protection pattern observed in the complex CIV indicates that $CopG_{WT}$ is bound mostly to the same face of the DNA, delimiting four consecutive regions of the major groove, wherein the protein could establish specific interactions with the DNA bases. This fact is consistent with the four binding subsites for CopG (LA, LSE, RSE and RA) in the operator DNA. Indeed, the two central regions match with those harboring the CopG tetramer bound to

the SE in the co-crystals (Gomis-Rüth, Solá *et al.* 1998a). However, once represented on a B-DNA double helix (10.5 bp/helical turn), some footprints seemed to comprise both faces of the DNA molecule (Figure 47A). For example, given that the footprinting *a'* was continued with *b'*, the footprint *a'*+*b'* was so large that spanned nearly a turn of the DNA helix.

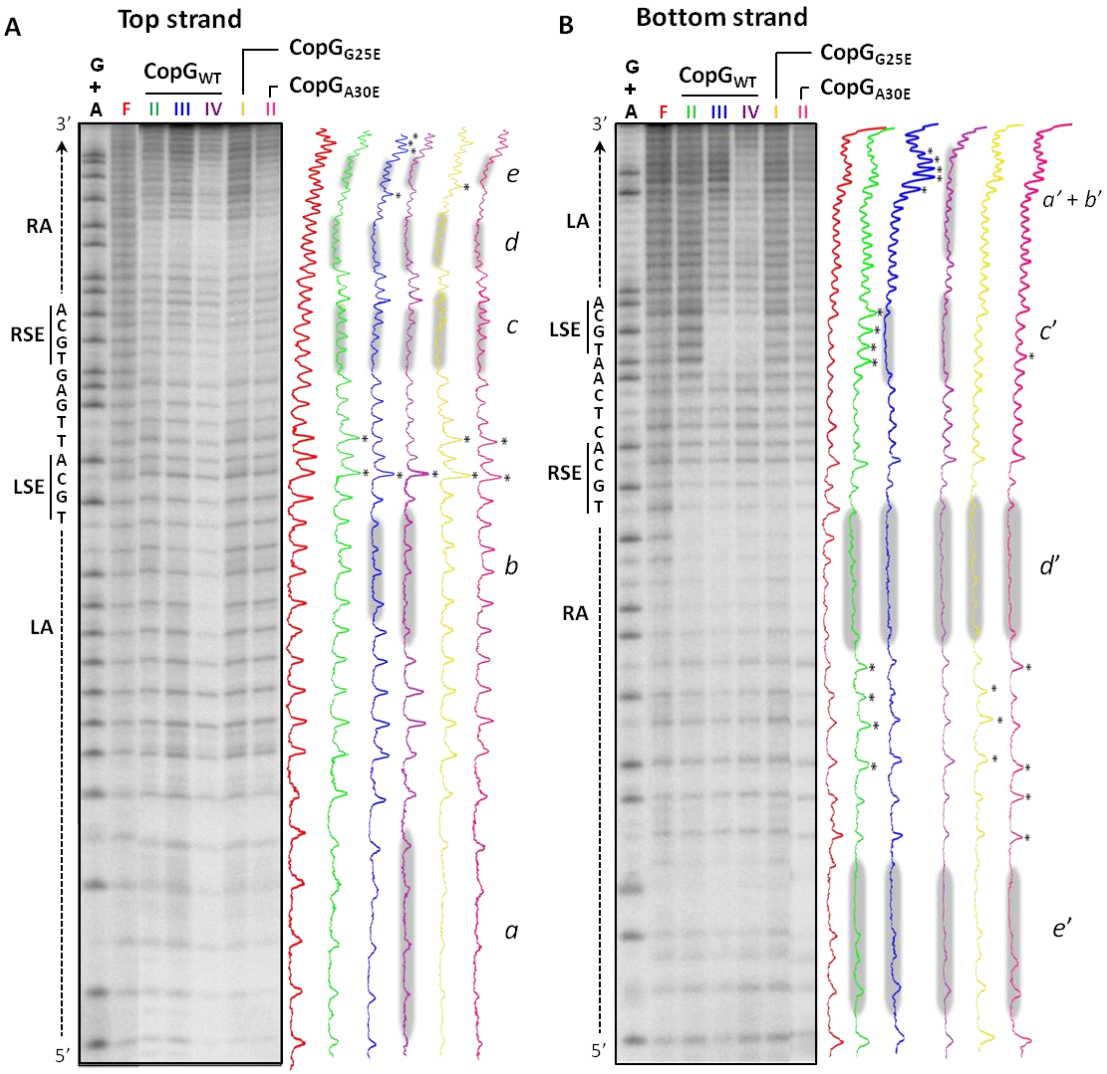


Figure 45. OH• protection footprinting of the wild-type and mutant CopG proteins bound to the wild-type operator DNA. The 55-bp WT oligonucleotide was 5'-end radiolabeled either at the top (A) or at the bottom (B) strand, and incubated with proteins for 20 min at 0 °C. After equilibrium, samples were exposed to OH• cleavage. Then, samples were isolated from PAA native gels and loaded onto sequencing gels. Lanes: F (*red*), free DNA; II (*green*), DNA of CopG$_{WT}$ complex CII; III (*blue*), DNA of CopG$_{WT}$ complex CIII; IV (*purple*), DNA of CopG$_{WT}$ complex CIV; I (*yellow*), DNA of CopG$_{G25E}$ complex CI; and II (*magenta*), DNA of CopG$_{A30E}$ complex CII. *G+A*, Maxam and Gilbert sequencing ladder for purines. Sequence of the 13-bp SE is shown, and repeats 5'-TGCA-3' in the LSE and RSE subsites are indicated. On the right, density-profiles of lanes are represented by the corresponding colors. *Gray-shaded* zones denote protections and *asterisks* hyper-exposition to cleavage. PAA 15% denaturing gels.

In the complex CIII, $CopG_{WT}$ protected a region spanning the coordinates 596 to 627 (Figure 47B). Four protected regions (*b*, *c*, *d* and *e*) were detected in the top strand, whereas in the bottom strand three regions were observed (*c*', *d*' and *e*'). Although the footprints *b* and *c*' were one nucleotide shorter than the corresponding ones in the complex CIV, the rest of protections were basically the same (Figures 45 and 47B). The analysis of protections in the complex CIII suggested that $CopG_{WT}$ was bound to the SE (LSE and RSE subsites) and to the RA. The most significant difference between the complex CIII and complex CIV involved the absence in the former of the protection corresponding to the LA. In contrast, a high reactivity of the sequence was observed between coordinates 585' and 589' (Figure 45 and 47B). Other positions hyper-exposed to the OH• attack were those at coordinates 623 and from 629 to 631.

The protections against the OH• cleavage in the complex CII formed by $CopG_{WT}$ spanned from coordinate 607 to coordinate 627 (Figure 47C). This protection pattern indicated that complex CII was generated by the binding of two protein dimers to the right half of the operator. Hence, footprints *c*, *d* and *e* were detected in the top strand, whereas in the bottom one, only the footprints *d*' and *e*' appeared. Protection in the region corresponding to the LSE was not observed; instead, the sequence spanning coordinates 600' to 603', as well as positions 601 and 603, were hyper-sensitive to the OH• attack (Figures 45 and 47C). The zone comprising the coordinates 617' to 620' in the bottom strand was also hyper-exposed to the OH• cleavage.

In the case of the complex CI generated by the mutant protein $CopG_{G25E}$, the protections were limited to a region spanning coordinates 607 to 619 (Figure 47D). In the top strand, two footprints were observed (*c* and *d*). The footprint *c* comprised an extra nucleotide towards the 3'-end, whereas the footprint *d* showed not variations compared to the rest of the complexes analyzed (Figure 47). In the case of the bottom strand, the only footprint detected was *d*'. Analysis of these protections suggested that $CopG_{G25E}$ interacts with the major groove in the region of the RSE. Hyper-expositions to the OH• cleavage were observed between positions 618' and 620' in the bottom strand, whereas positions 601, 603 and 623 appeared hyper-sensitive to the cleavage in the top strand.

On the other hand, the protections observed in the complex CII generated by the protein $CopG_{A30E}$ confirmed the binding of this mutant to the right half of the operator. These protections spanned from coordinate 607 to 627 (Figure 47E). As it can see, the protection pattern was basically the same as that observed in the complex CII formed by the wild-type protein (Figure 47C and E). However, the distinct pattern of hyper-expositions suggests that the DNA could acquire different conformations as a function of the protein bound. Positions 601 and

603 were hyper-exposed to the OH• attack in the top strand, whereas 603', 617' and from 620' to 622' appeared hyper-sensitive to the cleavage in the bottom strand.

7.2. Analysis of the interactions established by CopG proteins with the DNA bases

To identify base-specific contacts between proteins (both wild-type and mutants) and DNA, methylation protection footprinting was undertaken (Figure 46). DMS is an effective and widely used probe for sequence-specific protein-DNA interactions. DMS predominantly methylates the 7-nitrogen of guanine and the 3-nitrogen of adenine. Thus, reactivity with G-residues occurs in the major groove and with A-residues in the minor groove. In general, these methylated bases are subsequently converted to strand breaks by cleavage with piperidine (G >A) or NaOH (A>G) and displayed in denaturing gels. Our experiments were designed to identify mainly the G-residues contacted through the major groove. Unlike the OH• cleavage reactions, the DMS treatment was carried out after bound and unbound DNA were separated on a PAA native gel. Bands corresponding to complexes and unbound DNA were excised and put into Eppendorf tubes where the methylation reaction was performed. After the treatment with piperidine, samples were resolved in sequencing gels. Results from protection against DMS methylation correlated well with the OH• cleavage experiments. In addition, these results provided information about the contacts established by the proteins with the G-residues through the major groove (Figure 47).

In the complex CIV, $CopG_{WT}$ bound to the wild-type operator protected bases G590, G600, G609, G613 and G619 in the top strand, whereas bases G589', G601', G610' and G612' and G623' appeared protected in the bottom strand (Figure 46 and 47A). This pattern is consistent with one protein dimer being bound to each of the binding subsites (LA, LSE, RSE and RA) (Figure 47A). Thereby, protected bases G590 and G589' showed that in the complex CIV, a protein dimer interacts through the DNA major groove in the region comprising the LA. On the other hand, protected bases G600 and G601' showed that a protein dimer interacts through the major groove in the region corresponding to the LSE, whereas protected bases G609, G610' and G612' showed that another dimer of protein makes contacts through the DNA major groove in the region corresponding to the RSE. Finally, protected bases G619 and G623', confirmed the binding of a protein dimer through the DNA major groove to the RA. It is worth noting that protection of various adenines (from A584 to A586, A594, A595 and A611) through the DNA minor groove corroborated the OH• protection patterns observed in the top strand, as well as the protected adenines A599' and A603' corroborated those in the bottom strand (Figure 47A). At the 5'-limit of the OH• footprint *c* in the complex CIV, G607 appeared hyper-methylated. In

general, the enhanced methylation could be due to two reasons: because CopG altered the DNA topology, bending the DNA molecule and/or widening the major grooves, or because there was a hydrophobic region of the protein close to the DNA, which facilitated methylation.

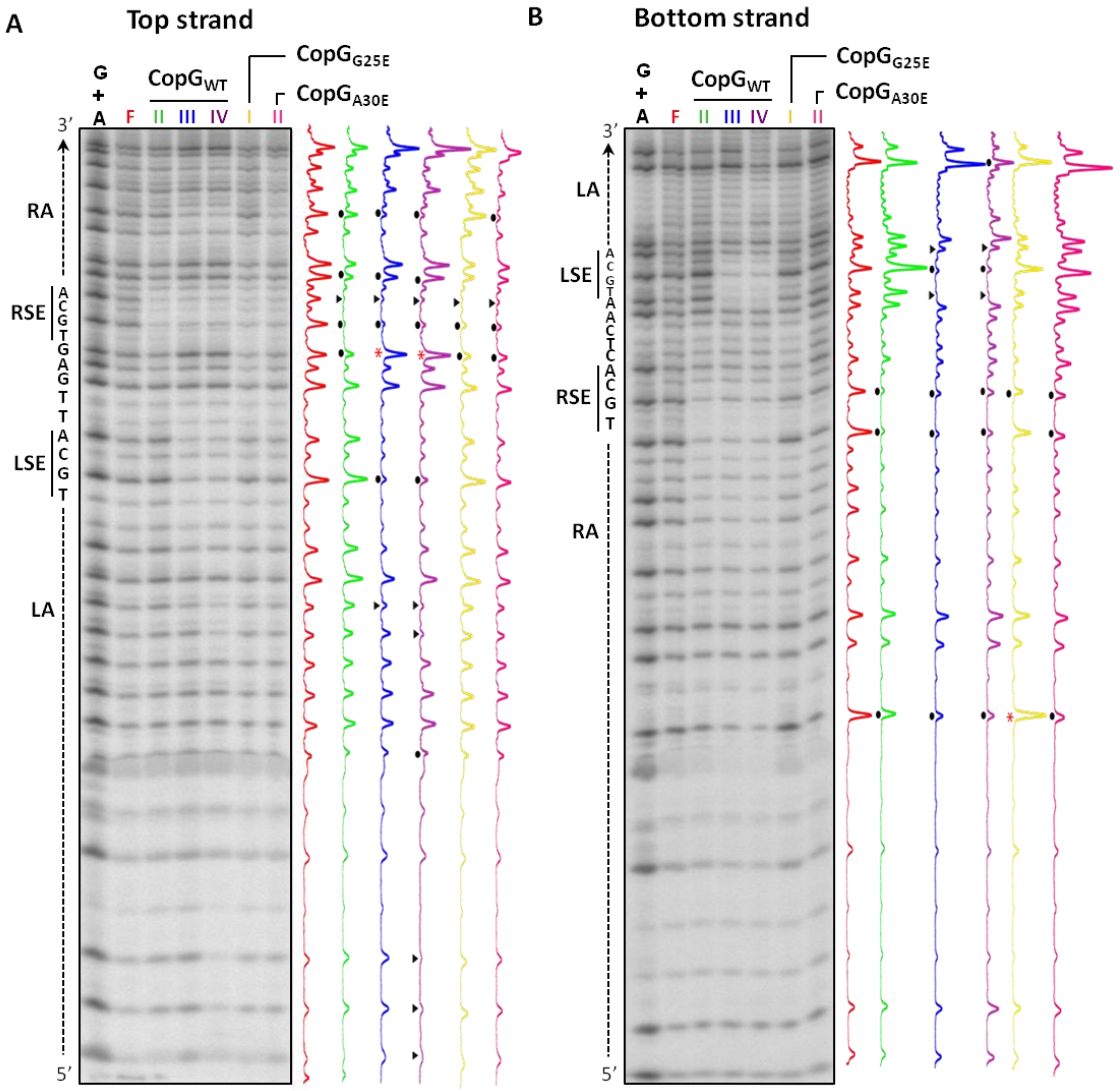


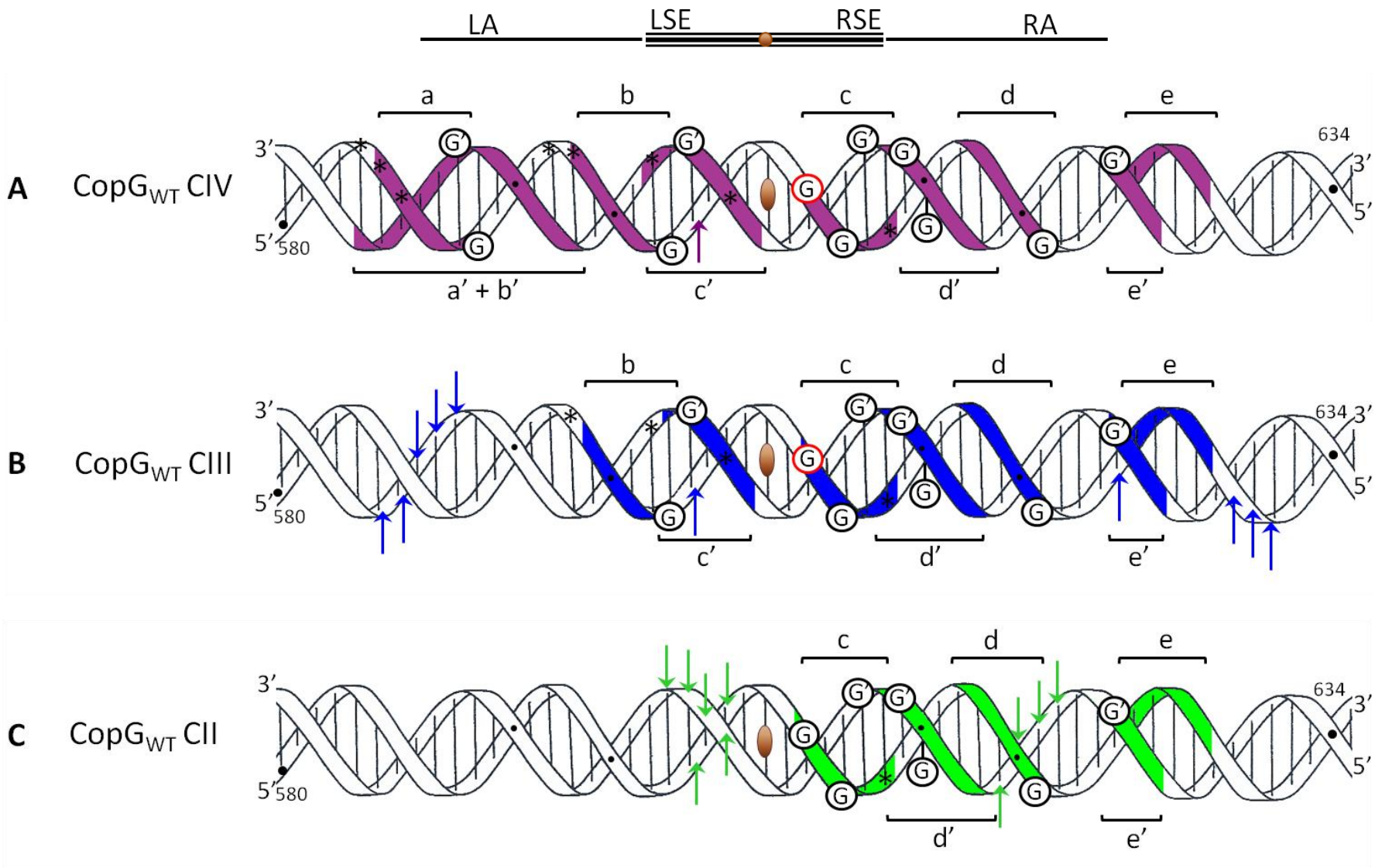
Figure 46. DMS-methylation protection footprinting of the wild-type and mutant CopG proteins bound to the wild-type operator DNA. The 55-bp WT oligonucleotide was 5'-end radiolabeled either at the top (A) or at the bottom (B) strand and incubated with proteins for 20 min at 0 °C. After equilibrium, samples were loaded onto PAA native gels, and the bands corresponding to the separate species of DNA were excised from them. Thereafter, free and complexed DNAs were exposed to DMS methylation, treated with piperidine and loaded onto sequencing gels. Lanes: F (*red*), free DNA; II (*green*), DNA of CopG$_{WT}$ complex CII; III (*blue*), DNA of CopG$_{WT}$ complex CIII; IV (*purple*), DNA of CopG$_{WT}$ complex CIV; I (*yellow*), DNA of CopG$_{G25E}$ complex CI; and II (*magenta*), DNA of CopG$_{A30E}$ complex CII. *G+A*, Maxam and Gilbert sequencing ladder for purines. Sequence of the 13-bp SE is shown and repeats 5'-TGCA-3' in the LSE and RSE subsites are indicated. On the right, density-profiles of lanes are represented by the corresponding colors. Small *black circles* and *red asterisks* on the left of each densitogram denote protection against and enhanced DMS-methylation of guanines, respectively. *Black triangles* denote the adenines protected against DMS methylation. PAA 15% denaturing gels.

Protected bases observed in the complex CIII (Figure 47B) were the same as those observed in the complex CIV in the regions corresponding to the central subsites and the RA, whereas, as expected, no evidence of binding to the LA was detected. Bases G590 and G589' did not appeared protected against methylation. Protected G600 and G601' at the LSE, as well as G609, G610' and G612' at the RSE, which were observed in complexes CIII and CIV, coincided with bases contacted by CopG in the co-crystals through the Arg4 residues from the protein β-sheet (Gomis-Rüth, Solá *et al.* 1998a). In the complex CIII, the protection of the adenine A595 in the top strand, and the bases A599' and A603' in the bottom one also confirmed the OH• protection patterns observed at the surroundings of the LSE (Figure 47B), whereas the protection of A611 corroborated the footprint *c* at the region of RSE.

Unlike the co-crystals, wherein two protein dimers were bound to the central subsites of the operator, the complex CII generated upon binding of $CopG_{WT}$ to the entire target DNA revealed the binding of two dimers to the right half of the operator, consisting of the RSE and RA subsites. Thus, the bases G619 and G623' located at the RA appeared protected as in the higher-order complexes (Figure 47C). In contrast, the hyper-exposed G607 of the RSE in the complexes CIII and CIV appeared protected in complex CII (Figure 47C). This difference in the protection pattern could suggest variations in the three-dimensional configuration of the dimer bound to the RSE depending on the complex formed, hence affecting the interactions between the protein and DNA.

The analysis of the methylation protection pattern in the complex CI generated by $CopG_{G25E}$ showed binding of a protein dimer to the RSE (Figure 47D). Bases G607 and G609 in the top strand, and G610' in the bottom one appeared protected. Base G612' was contacted by the wild-type protein in complexes CII, CIII and CIV but not by $CopG_{G25E}$ in the complex CI. As it occurred with the rest of the complexes, the adenine A611 also appeared protected. The absence of protections in the LA, LSE or RA regions indicates that this mutant does not interact with DNA in these subsites.

Finally, the protections against DMS methylation observed in the complex CII generated by protein $CopG_{A30E}$ were almost identical to the protections observed in the complex CII with $CopG_{WT}$ in the right half of the operator (Figure 47E). The only difference was that the base G613 was not protected by the mutant protein in the top strand. These results confirmed that mutant $CopG_{A30E}$ only binds to the right half of the operator when bound to the wild type DNA.


LA
LSE
RSE
RA
A
CopG$_{WT}$ CIV
B
CopG$_{WT}$ CIII
C
CopG$_{WT}$ CII
a
b
c
d
e
3′
5′
580
634
a′ + b′
c′
d′
e′

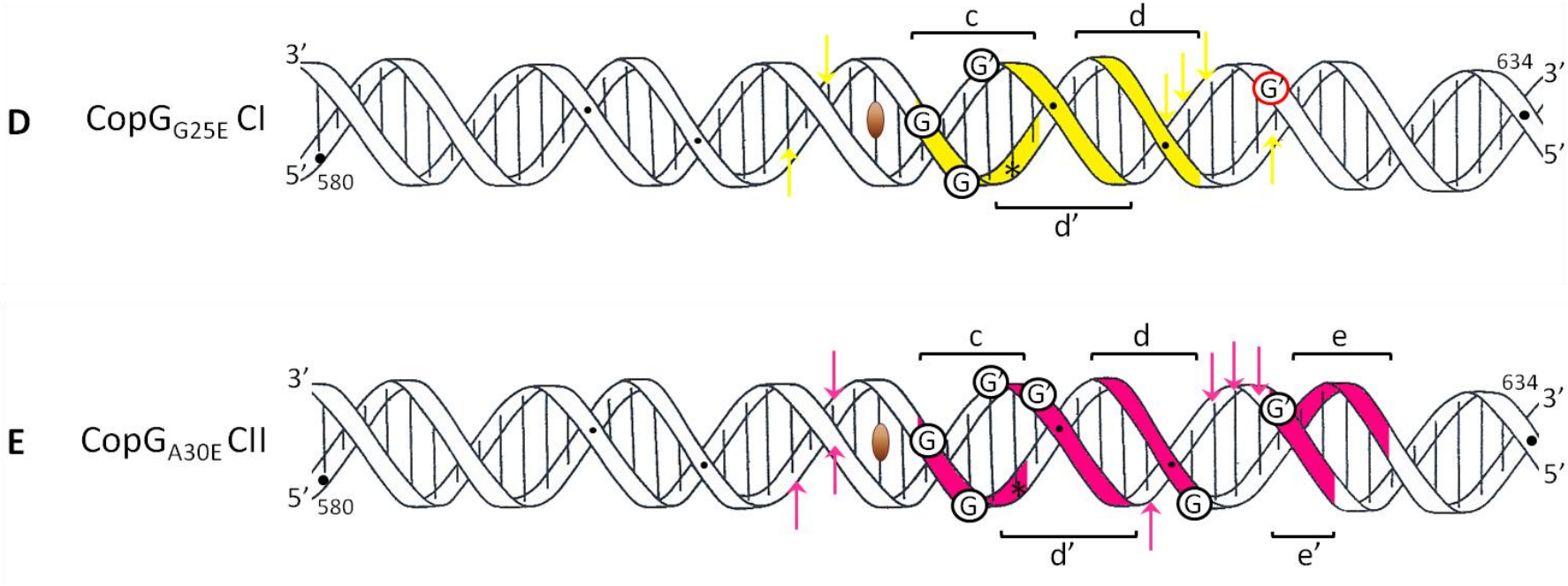


Figure 47. Summary of footprinting assays of the different complexes generated by the three protein variants bound to the wild-type operator DNA. (A, B, C) Characterization of complexes CIV (*purple*), CIII (*blue*) and CII (*green*) generated by CopG$_{WT}$ bound to the DNA operator, respectively. (D) Characterization of complex CI (*yellow*) formed by CopG$_{G25E}$. (E) Characterization of complex CII (*magenta*) generated by CopG$_{A30E}$. Protein-DNA interactions are depicted on a double-helix of B-DNA (10.5 bp/helical turn). Colored regions on the DNA backbone denote the protections against the OH• cleavage, which are also indicated by *brackets up* and *down* of the schemes. *Colored arrows* indicate sites hyper-sensitives to OH• attack. Protected and hyper-methylated G-residues are *black-* or *red-encircled*, respectively. The two-fold symmetry axis of the SE is indicated by a *brown ellipse*. A schematic representation of the operator is shown on top of the figure; the two-fold symmetry axis of the SE is indicated by a *small circle*. Coordinates of the pMV158 sequence are indicated on the double helix. *Asterisks* indicate the A-residues protected through the DNA minor groove.

Along this chapter, the interactions made by mutant and wild-type proteins with the operator DNA have been described in detail. It has been shown that the deficiencies in binding affinity and cooperative properties of $CopG_{G25E}$ and $CopG_{A30E}$ depend both on their own features and on the nucleotide sequence and arrangement of the subsites within the operator. On the other hand, the selective binding of the mutant $CopG_{A30E}$ to the right operator half, as well as the distinct patterns of contacted bases at the RSE in the complexes CII (with $CopG_{WT}$ and $CopG_{A30E}$) compared to those at the primary site but in the complexes CIII and CIV, could suggest the existence of an interaction surface between the protein dimers that differs from that described in the co-crystals (see *Discussion*). This new putative interface would be established between the central and lateral subsites and might have a great influence on the cooperative and sequential binding of the protein to the secondary subsites of its operator DNA.

CHAPTER IV

Super-operators

Another strategy to address the study of the cooperativity and affinity properties of CopG was probing the effect of optimizing the binding sites. Before starting this thesis, a new set of oligonucleotides were designed with the aim of increasing their affinity of binding to CopG relative to the WT operator, and hence, were properly named *super-operators* (Figure 48, Table 7). The protein binding subsites within the operator were modified in order to create sites similar to the RSE. Therefore, the super-operators conserve the intact RSE, and have an extra RSE-like binding site. The following data were considered for modifying the sequence in the super-operators: i) the fact that the RSE is the primary site of binding of CopG (*Chapter I*) (Hernández-Arriaga, Rubio-Lepe *et al.* 2009) ii) the protein-DNA interactions observed in the co-crystals obtained from two CopG dimers bound to the 19-bp SE (Gomis-Rüth, Solá *et al.* 1998a), and, iii) the protections observed in preliminary studies of OH• footprintings (del Solar, Pérez-Martín *et al.* 1990) and DMS footprintings carried out at room temperature (unpublished results). In the WT operator, four 5'-TGCAC-3' (or variant) recognition sequences spaced by 5 nt were defined as being the binding sites. Thus, the sequences recognized by CopG in the SE were those described in the co-crystals (5'-TGCAC-3' for the RSE, 5'-TGCAA-3' for the LSE), whereas the proposed recognition sequences in the lateral subsites were 5'-TTCGT-3' and 5'-TGCTA-3' in the LA and the RA, respectively (top of Figure 48).

On the other hand, although the CopG dimer has local two-fold symmetry, in the co-crystals, neither the interactions established by $CopG_{WT}$ in each subsite of the SE nor the bases contacted in these subsites are symmetrical. This allows the assignment of the relative orientation of the proposed recognition sequences at the different subsites (Figure 48). Hence, it was interesting to test whether the relative orientation of the new 5'-TGCAC-3' sequence introduced within the super-operators affected the affinity of binding by CopG. Differences among the super-operators harboring the same improved subsite consisted in: i) variations in the bases surrounding the new 5'-TGCAC-3' sequence, ii) alteration of the relative orientation of this sequence, and iii) variations in the distance from the new RSE-like site to adjacent subsites. Assays carried out using super-operators gave us further details about specificity of protein binding to its operator. The relative affinity of the wild-type protein for each super-operator, as well as the effect of the different changes on binding of mutant CopG proteins was analyzed, and the results are described all along this chapter.

1. Relative binding affinity of $CopG_{WT}$ to the super-operators

The affinity of binding of CopG to the super-operators was calculated by competitive EMSAs (Figure 49). For this, CopG was incubated with a mixture of two DNAs, one of them being any of

the 55-bp super-operators, and the other, the 239-bp DNA fragment containing the wild-type operator. Bound and unbound DNAs were resolved in PAA native gels and analyzed as described in *M&M* (Part II, 2.1). Results are summarized in Table 25.

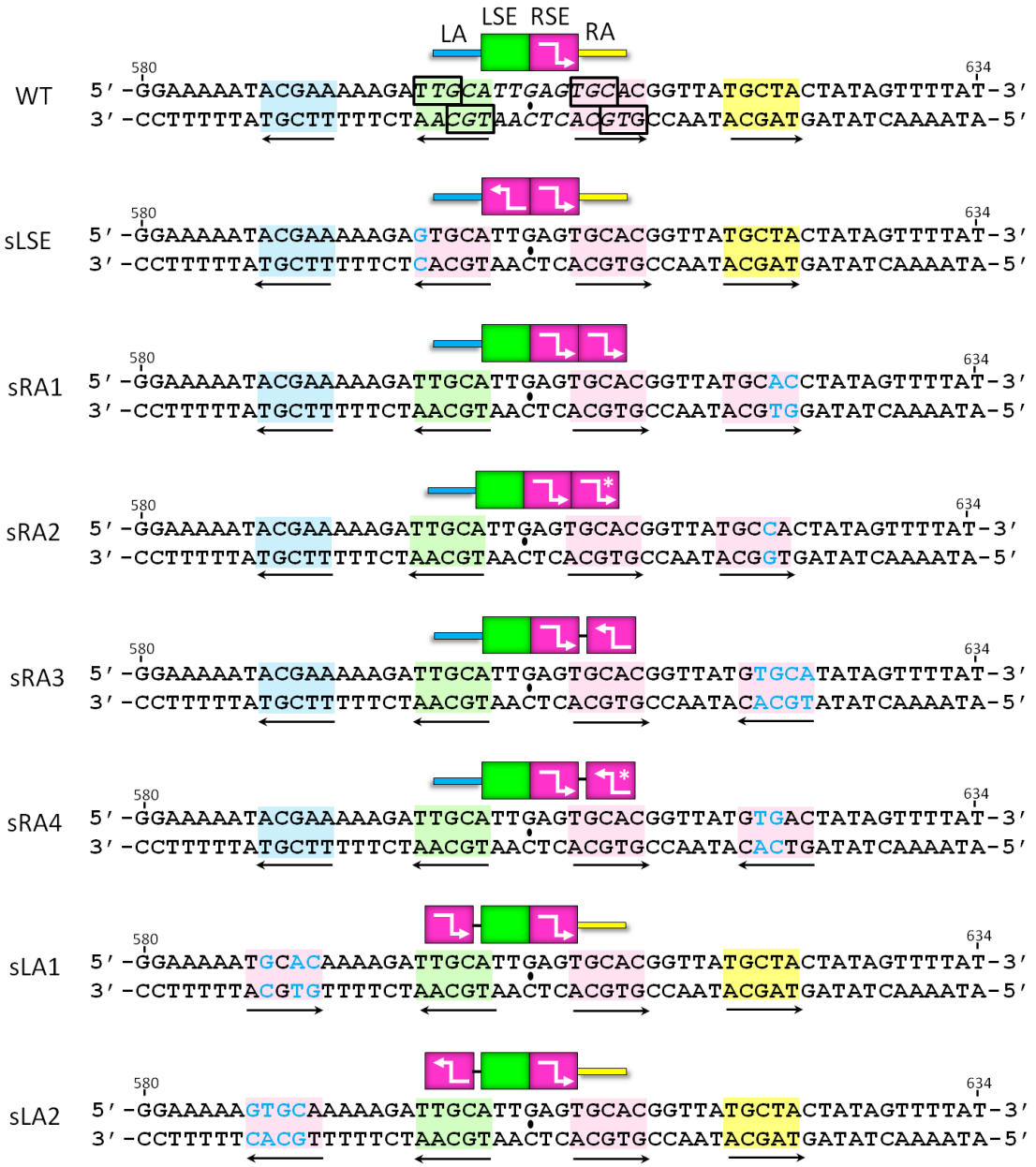


Figure 48. **Schemes and nucleotide sequences of the super-operators.** The 13-bp SE is displayed in *italics* in the WT operator with its two-fold symmetry axis indicated by a *small circle*. Subsites in the SE are *green* (LSE) and *pink* (RSE) *shaded*. Based contacted by $CopG_{WT}$ in the LSE and RSE (as observed in 19-bp SE-protein co-crystals) are *boxed* in the WT operator. New RSE-like binding sites are also *pink shaded*. Bases that were expected to interact with protein in the LA and RA subsites are *blue* and *yellow shaded* respectively. Nucleotides modified in the mutant operators are in *blue*. *Arrows* indicate the orientation of the sequence 5'-TGCAC-3' (or variants). *Asterisks* indicate variations in the sequence 5'-TGCAC-3'. Variations in the inter-subsite spacing relative to the WT operator are denoted by a *small line* between them in the diagrams.

As can be observed, the sLSE super-operator exhibited the highest affinity for CopG, with an increase of near 18-fold relative to the wild-type operator DNA (Table 25, Figure 49). Changing the thymine (598) to guanine in the LSE of the operator likely results in new strong interactions between the G:C base-pair and the Arg4 residues of the protein, thus increasing the binding affinity. Regarding the other super-operators, only sRA3 and sLA1 showed a moderate increment (6- and 4-fold, respectively) in the binding affinity of $CopG_{WT}$, whereas the changes in the rest of super-operators did not have any effect on binding of protein (Table 25, Figure 49).

Table 25 | $CopG_{WT}$ binding affinity to different super-operators.

Super-operator		Relative affinity[a]
WT		1
sLSE		18.67 ± 5.49
sRA1		1.08 ± 0.14
sRA2		0.73 ± 0.1
sRA3		6.15 ± 1.54
sRA4		0.74 ± 0.1
sLA1		3.91 ± 0.73
sLA2		0.82 ± 0.13

[a] The relative affinities calculated for the different operator variants were further normalized with respect to that obtained for the WT oligonucleotide (0.46 ± 0.06). In diagrams, *colored boxes with filled lines* represent the conserved subsites (*blue* for LA, *green* for LSE, *magenta* for RSE, and *yellow* for RA). The orientation of the RSE 5'-TGCAC-3' sequence is drawn by *arrows*. Variations in the inter-subsite spacing relative to the WT operator are denoted by a *small line* between them, and variations in the sequence of the new RSE-like sites are marked by *asterisks*.

The complex-formation patterns of $CopG_{WT}$ bound to the super-operators exhibited differences compared to the WT operator. In the cases wherein the sequence modifications improved the binding of the protein to the target DNA (sLSE, sRA3 and sLA1) (Figure 49), the most populated complexes might result from a 'better' binding of the protein to the 'improved' subsite, hence reflecting the sequential order of occupancy of the operator subsites. Thus, the higher proportion of complex CIII generated by binding of the protein to the sLSE super-operator, as compared to the WT operator, would indicate that once the LSE subsite (which accommodates the third CopG dimer) was improved, the formation of complex CIII was facilitated, albeit the binding affinities for other subsites and overall cooperativity remained unmodified. Alike, the higher proportion of complex CIV observed with the sLA1 super-operator might not indicate an

increase of cooperativity, but the increase in affinity of binding to the last subsite to be bound, the LA, which was improved in this super-operator.

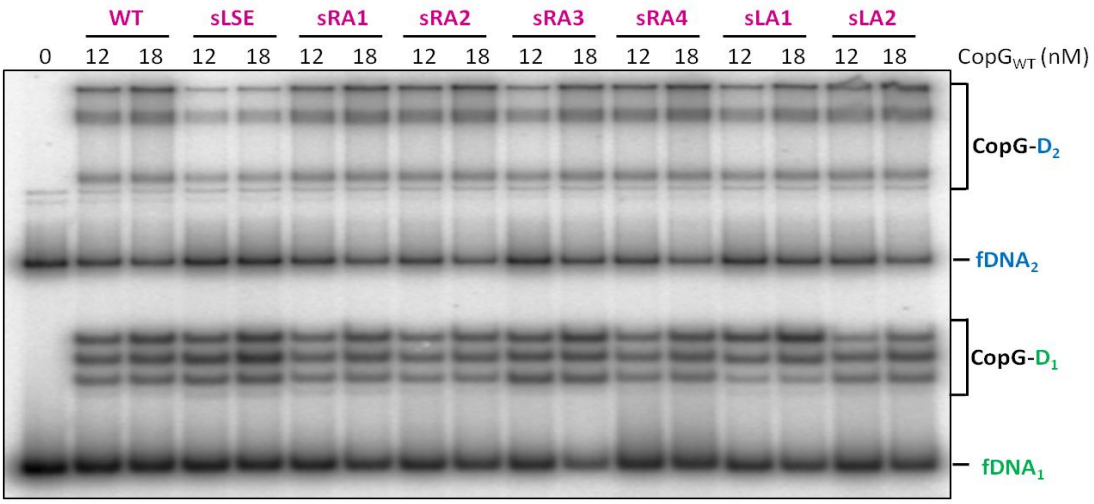


Figure 49. **Competitive EMSAs to determine the relative binding affinity of CopG$_{WT}$ to the super-operators**. CopG (at the stated concentrations) was incubated at room temperature with a mixture of the indicated 55-bp super-operator (*DNA$_1$*/*D$_1$*) and the 239-bp wild-type operator DNA fragment (*DNA$_2$*/*D$_2$*). Complexes generated by binding of CopG to each DNA (*CopG-D$_1$*/*CopG-D$_2$*) and unbound DNA (*fDNA*) are indicated.

2. Binding of mutant CopG proteins to the super-operators

Although the relative affinity of mutant CopG proteins for the super-operators was not measured, it was interesting to disclose the complex-formation patterns generated by them (Figures 50 and 51). In general, the qualitative binding of these mutants to the super-operators correlated with the affinities measured for the native protein. Target DNAs that behaved as super-operators for CopG$_{WT}$ (sLSE, sRA3 and sLA1) seemed to do the same also for the mutant proteins.

In the case of CopG$_{G25E}$, only one complex migrating similarly to the complex CI formed by CopG$_{WT}$ was observed at room temperature with any of the super-operators (Figure 50A). On the other hand, as it occured with the WT operator, in experiments carried out at 0 °C, the pattern of binding accomplished by CopG$_{G25E}$ showed the non-cooperative appearance of a second and even a third complex with most of super-operators. Only two of the super-operators assayed with this mutant at 0 °C are shown in Figure 50B. The difference between the complexes that this mutant protein formed with the WT operator and with the super-operators lies mainly in the intensity of the bands corresponding to the complexes. Considering that CopG$_{G25E}$ only binds to the primary site, and that super-operators are expected to contain two primary sites, the

complex CI may consist of a mixture of complexes with a protein dimer bound to different subsites. This could also explain the higher proportions of complex CI observed with some of the super-operators (sLSE, sRA3 and sLA1) obtained at room temperature (Figure 50A). Preliminary results from OH• footprinting experiments performed by isolating the treated complexes from PAA native gels supported this assumption (not shown). Actually, the complex CI generated by binding of $CopG_{G25E}$ to the sLSE and sRA3 super-operators at 0 °C exhibited protections in the regions that corresponded to the RSE and to the new RSE-like binding site. Therefore, we could say that the band representing the complex CI was composed by the sum of the different possible complexes consisting of a protein dimer bound to either "primary site" (Figure 50A). In addition, under the same experimental conditions, the protections observed in the complexes CII generated by $CopG_{G25E}$ bound to the sLSE and sRA3 super-operators showed that two dimers bound to both 'primary sites' of the target DNAs (not shown).

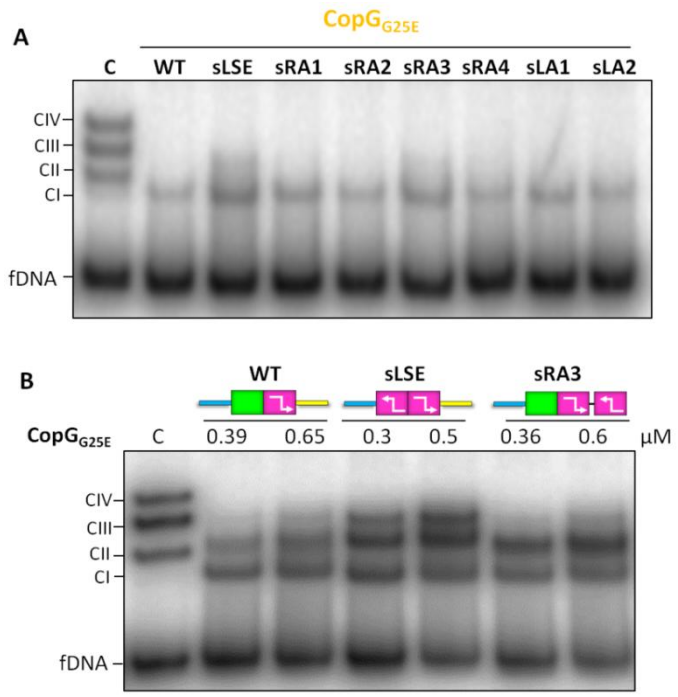


Figure 50. Binding of $CopG_{G25E}$ to the super-operators. (A) $CopG_{G25E}$ was mixed with the radiolabeled super-operators (9-14 nM, Table 12) at a molar ratio of 1:20 (DNA:protein protomers) and incubated at room temperature in the presence of T7-bacteriophage DNA as competitor. (B) Pattern of complexes generated by binding of $CopG_{G25E}$ to the sLSE and sRA3 super-operators. The protein (at the stated concentrations) was incubated with radiolabeled WT (13 nM), sLSE (11 nM) or sRA3 (12 nM) at 0 °C in the presence of 120 µg/ml of heparin. In both panels, the complex-formation pattern of $CopG_{WT}$ bound to the WT operator is shown as reference (lane C). Bound and unbound DNAs are indicated. In diagrams, *colored boxes with filled lines* represent the operator subsites (*blue* for LA, *green* for LSE, *magenta* for RSE, and *yellow* for RA). *Arrows* indicate the orientation of the RSE 5'-TGCAC-3' sequence, which is contacted by $CopG_{WT}$ in the co-crystals with the SE. Variations in the inter-subsite spacing are denoted by a *small line* between them.

At room temperature, this mutant could bind to either subsite but not to both, while at low temperature it generated a second complex. This indicates that at room temperature, a hindrance between the dimers bound to adjacent subsites exists, and that the binding of one of them prevents the binding of the adjacent dimer, hence exhibiting negative cooperativity. On the other hand, at low temperature, the higher twist of DNA (less than 10.5 bp/helical turn) could alleviate this situation as both subsites would be rotated relative to each other on the DNA double helix. This would decrease the interference between adjacent dimers, allowing the formation of complex CII, but without apparent positive cooperativity. The variations in migration of the $CopG_{G25E}$-complex CII compared to the $CopG_{WT}$-complex CII are likely due to a distinct effect of protein binding on bending of the DNA molecule. Be it as it may, on this theme more experiments are needed for drawing conclusions. Overall, these results confirmed again that protein $CopG_{G25E}$ is defective in cooperativity: it was able to bind to the RSE and to the improved (RSE-like) subsites because of its intrinsic affinity for the primary site, but no protein-protein interactions seemed to be feasible.

On the other hand, binding of the $CopG_{A30E}$ mutant protein to the super-operators was a bit different. Unlike $CopG_{WT}$ and $CopG_{G25E}$, the only super-operator for which this mutant seemed to show increased binding affinity was the sRA3 (Figure 51A). This fact is consistent with the stable binding of $CopG_{A30E}$ as a tetramer only to the right half of the operator (*Chapter III*, Figures 37, 39A and 47E). However, we expected that this mutant could generate a complex CIV with those super-operators optimized in the left half (namely, sLSE and sLA1), but this did not occur (Figure 51A). In EMSAs carried out at room temperature, $CopG_{A30E}$ was only able to form complexes migrating like the complex CII generated by $CopG_{WT}$ bound to the WT operator (Figure 51A).

At low temperature (0 °C), $CopG_{A30E}$ generated the complex CII (a tetramer at the right operator half) when bound to the WT operator (not shown). Under the same conditions, a great smear appeared over the expected specific band (the CII-like complex) generated by $CopG_{A30E}$ bound to the sLSE super-operator (Figure 51B). This could suggest that once the complex CII was formed, another dimer was attempting to occupy the improved (RSE-like) LSE. However, as far as we know, the mutant $CopG_{A30E}$ has a very low affinity for binding to DNA as a single dimer, so that we only detect specific complexes if the protein is bound as a tetramer. Therefore, the unclear slowest-migrating band that appeared in the gel could correspond to a nucleoprotein complex carrying a tetramer bound to each operator half, but with a configuration and hence, an electrophoretic mobility, different from those acquired by binding of two $CopG_{WT}$ tetramers to the WT operator (Figure 51B). However, according to the migration pattern, it also could

correspond to a complex formed by three dimers bound (to the RSE, RA and improved LSE subsites). Thus, although affinity of a single $CopG_{A30E}$ dimer for DNA is very low, the protein could be accommodated somehow in the improved LSE and establish contacts with the RSE-bound dimer. As explained for mutant $CopG_{G25E}$, this could be due to the overtwisting of the DNA caused by the low temperature. This change in DNA topology would rotate the adjacent subsites on the double helix, thus promoting interactions for positive cooperativity (or at least, avoiding negative cooperativity). Even so, this putative nucleoprotein complex seems to be too unstable to form a defined band (Figure 51B). In the case of sRA3 at low temperature, the smear was much less notable because, in this super-operator, the modified sequence 'improved' the binding of $CopG_{A30E}$ only to the right operator half. Unlike the sLSE super-operator, in sRA3, a new RSE-like subsite promoting the binding of $CopG_{A30E}$ to the left operator half was not created (Figure 51B).

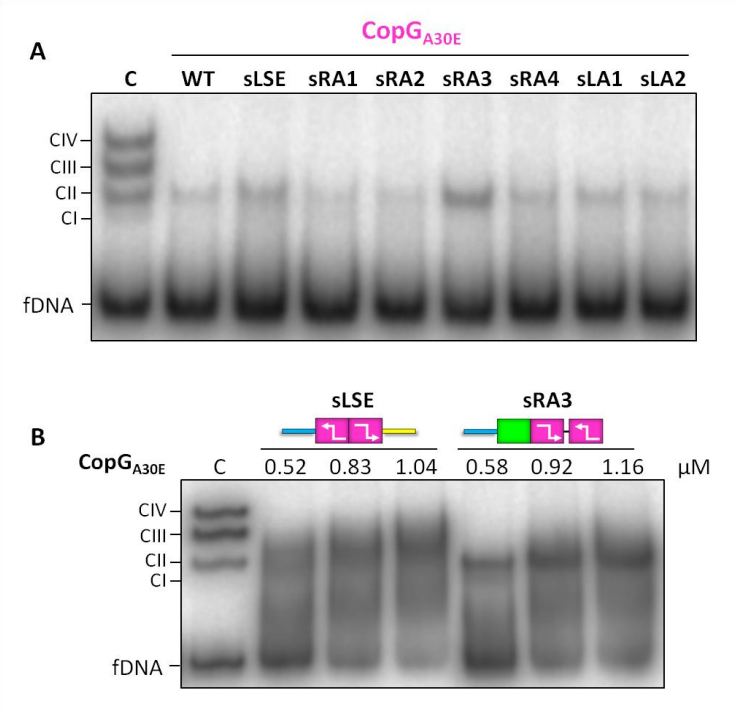


Figure 51. Binding of $CopG_{A30E}$ to the super-operators. (A) $CopG_{A30E}$ was mixed with the radiolabeled super-operators (9-14 nM, Table 12) at a molar ratio of 1:50 (DNA:protein protomers) and incubated at room temperature in the presence of T7-bacteriophage DNA as competitor. (B) Pattern of complexes generated by binding of $CopG_{A30E}$ to the sLSE and sRA3 super-operators. The protein (at the stated concentrations) was incubated with radiolabeled sLSE (11 nM) or sRA3 (12 nM) at 0 °C in the presence of 120 µg/ml of heparin. In both panels, the complex-pattern of $CopG_{WT}$ bound to the WT operator is shown as reference (lane C). Bound and unbound DNAs are indicated. In diagrams, *colored boxes with filled lines* represent the operator subsites (*blue* for LA, *green* for LSE, *magenta* for RSE, and *yellow* for RA). *Arrows* indicate the orientation of the RSE 5'-TGCAC-3' sequence, which is contacted by $CopG_{WT}$ in the co-crystals with the SE. Variations in the inter-subsite spacing are denoted by a *small line* between them.

These results suggest that, at room temperature, the incapability of mutant $CopG_{A30E}$ to bind to both central subsites in the operator is more likely due to an inappropriate dimer-dimer interaction surface than to a lower binding affinity to the sequence of the LSE relative to that of the RA, since "improving" the LSE, although greatly increased the binding of $CopG_{WT}$ to the SE, failed to overcome the binding defect of $CopG_{A30E}$ (see below).

3. Effect of altering the sequence of the operator subsites or the distance between them on protein binding

The analysis of the affinities of the wild-type protein for the different super-operators, as well as the complex-formation patterns observed with mutants, demonstrated that the nucleotide-sequence of the subsites is not entirely responsible for the binding of CopG to its operator DNA. Therefore, spacing between the subsites could also affect the binding affinity of the protein. To address this issue, a new set of oligonucleotides comprising exclusively two adjacent subsites with varying nucleotide sequence and spacing were synthesized. Some of the modifications made in the super-operators were also included in these two-subsite operators (see *M&M,* Table 7 for description). Changes in the distance between nearby subsites were performed either by inserting or by deleting a single base-pair (Figure 52). Binding of proteins $CopG_{WT}$ and $CopG_{A30E}$ to these new constructions was analyzed by EMSA (Figure 53).

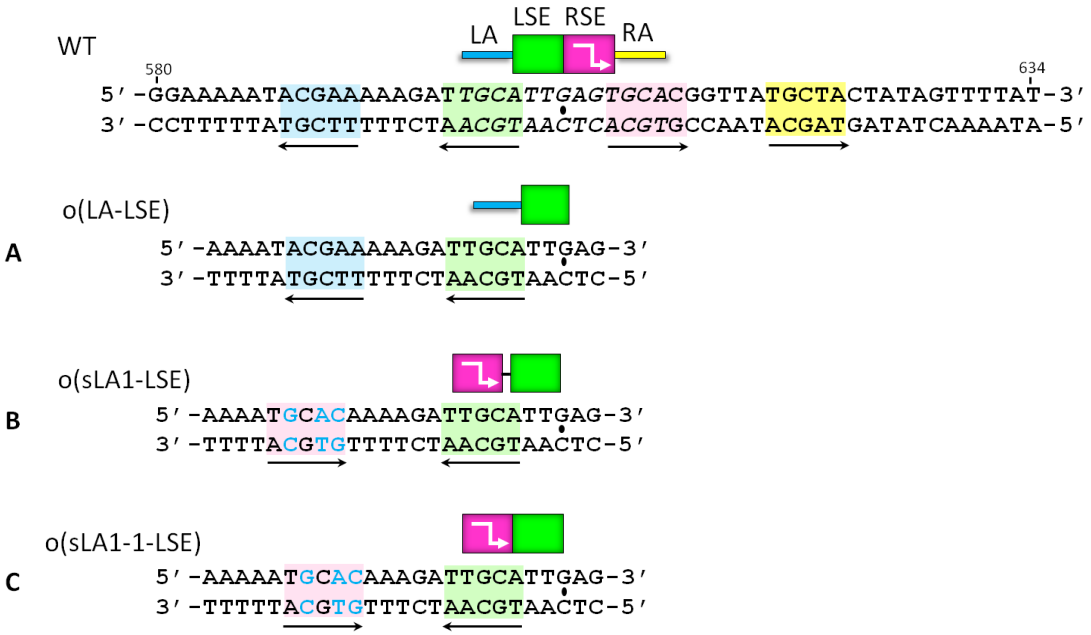


Figure 52. **Schemes and nucleotide sequences of different super-operator variants consisting only of two subsites.** The 13-bp SE is displayed in *italics* in the WT operator with its two-fold symmetry axis indicated by a *small circle*. Bases contacted by $CopG_{WT}$ (observed in co-crystals) in the LSE and RSE are *green* and *pink shaded* respectively. Bases that are likely to interact with protein in the RA and LA subsites are *blue* and *yellow shaded* respectively. Optimized RSE-like binding sites are also *pink shaded*. Nucleotides modified in the mutant operators are in *blue*. *Arrows* indicate the orientation of the RSE 5'-TGCAC-3' sequence, which was contacted by $CopG_{WT}$ in the co-crystals with the SE. Variations in the inter-subsite spacing are denoted by a *small line* between them.

Under the experimental conditions employed, $CopG_{WT}$ was hardly able to form a complex with the o(LA-LSE) operator (Figure 53A). This result is consistent with the fact that the left half of the operator has the lowest contribution to the overall affinity of the protein (*Chapter III,* Table 23).

However, $CopG_{WT}$ bound with higher affinity to the o(sLA1-LSE) operator, wherein the distance between the 'improved' LA and the LSE subsites is 1-nt larger than the distance between the equivalent subsites in the oligonucleotide carrying the intact left operator half, the o(LA-LSE) operator (Figure 53A, lane B). This behavior, although accompanied with a lower affinity, was also observed with mutant $CopG_{A30E}$. Mutant $CopG_{A30E}$ was unable to bind to the o(LA-LSE) operator consisting exclusively of the LA and LSE subsites, as observed also with the full-length operator LA-LSE (Figure 37); however, it did bind to the o(sLA1-LSE) operator (Figure 53B, lanes A and B). This result indicates that mutant $CopG_{A30E}$ can become able to bind to the left half of the operator provided that the sequence of the LA element is optimized and the spacing between subsites is enlarged as compared to the WT (Figure 53A and B). In previous experiments, $CopG_{A30E}$ was shown to bind to the full-length sLA1 super-operator; however, we have not analyzed whether the observed complex CII was indistinctly populated by complexes formed by tetramers of the mutant protein bound to either operator halves. In addition, specific complexes of higher order than CII were not detected, indicating that tetramers of $CopG_{A30E}$ bound to the right or left half of the operator hinder each other. The positive effect on $CopG_{WT}$ binding affinity to the o(sLA1-LSE) operator was reversed by shortening (-1 bp) the spacer DNA between the improved LA and LSE subsites (the o(sLA1-1-LSE) operator), thus returning to the wild-type spacing of 5 nt. Moreover, although formation of the complex CII with this super-operator exhibited an apparently higher affinity compared to the o(LA-LSE), it was lower compared to the o(sLA1-LSE) variant (Figure 53A, lanes A-C). In the case of the $CopG_{A30E}$ mutant, no band shifts were observed with this impaired DNA, probably because of its diminished affinity for DNA (Figure 53B, lane C).

It is important to highlight the different migration of the unique complex observed by binding of $CopG_{WT}$ to the o(LA-LSE) super-operator (Figure 53B). We assumed that it might correspond to a tetramer bound, since formation of a stable nucleoprotein complex in the left half of the operator requires the simultaneous binding of two CopG dimers, yielding the complex CII and giving an appearance of huge cooperativity. This complex CII-formation pattern has been previously observed (Figure 42) and also for the 55-bp LA-LSE operator (Figure 41). The different electrophoretic migration of the complex formed on the o(LA-LSE) operator could be due to a distinct configuration of the nucleoprotein complex. In fact, faint differences in the electrophoretic mobility of the complexes CII formed on the 55-bp operators had already been observed, depending on the operator half containing the tetramer bound (Figure 40).

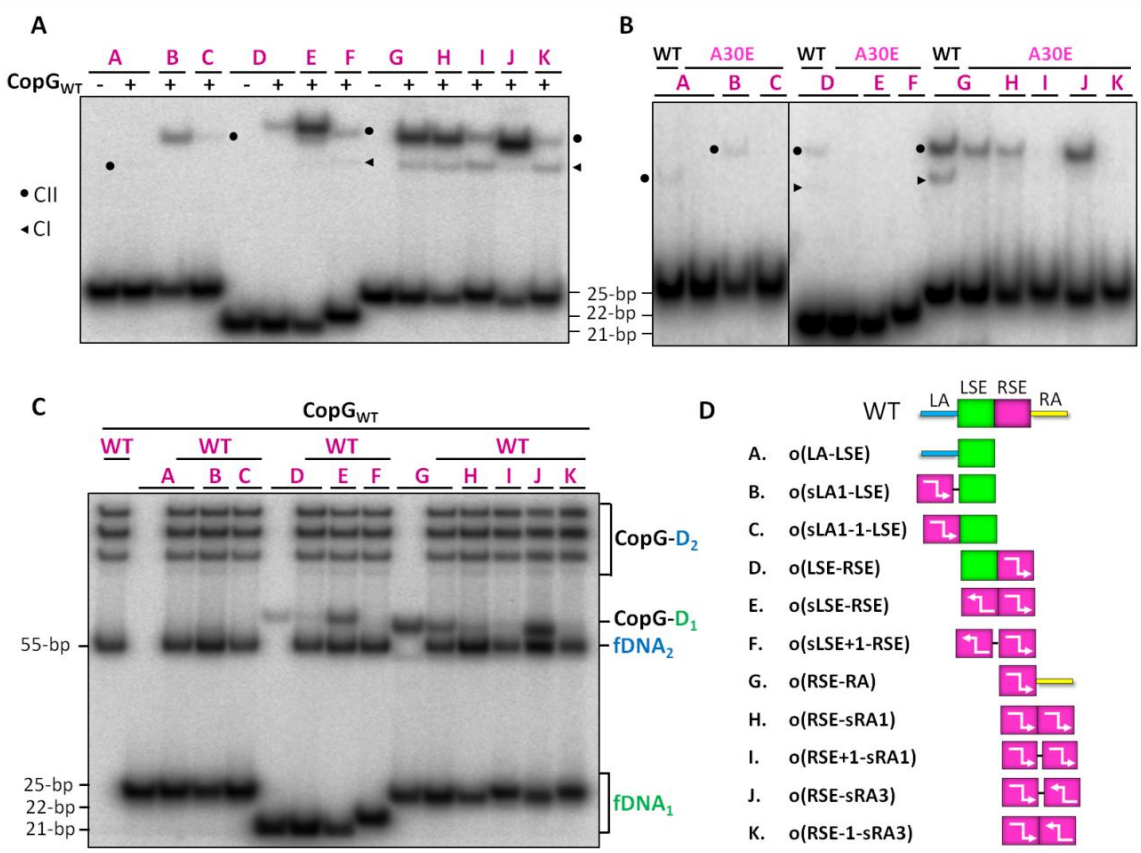


Figure 53. **Binding affinity of CopG$_{WT}$ and CopG$_{A30E}$ to different super-operator variants consisting only of two subsites**. Proteins CopG$_{WT}$ (*A*) and CopG$_{A30E}$ (*B*) were incubated at room temperature with radiolabeled DNAs (3-9 nM, Table 13), at a molar ratio of 1:3 and 1:6 DNA:protein protomers, respectively. DNA of T7-bacteriophage was added as competitor. *Circles*: CII complexes; *Triangles*: CI complexes. PAA 10% (*A*) and PAA 15% (*B*). (*C*) Competitive EMSAs, PAA 10%. CopG$_{WT}$ (at a molar ratio of 1:6 DNA:protein protomers) was incubated at room temperature, with a mixture of the indicated operator variant comprising exclusively two subsites (*DNA$_1$/D$_1$*), and the WT operator (*DNA$_2$/D$_2$*). Complexes generated by binding of CopG to each DNA (*CopG-D$_1$/CopG-D$_2$*) and unbound DNA (*fDNA*) are indicated. (D) Diagrams of the operator variants consisting of only two subsites. *Colored boxes with filled lines* represent the conserved subsites (*blue* for LA, *green* for LSE, *magenta* for RSE, and *yellow* for RA). The orientation of the RSE 5'-TGCAC-3' sequence, which was contacted by CopG$_{WT}$ in the co-crystals with the SE, is drawn by *arrows*. The scheme of the WT operator is shown as a reference. Variations in the inter-subsite spacing are denoted by a *small line* between them.

Using the operator variants that comprise exclusively the SE, the higher affinity of CopG$_{WT}$ for the o(sLSE-RSE) than for the o(LSE-RSE) operator was demonstrated by a higher fraction of complex CII (Figure 53A, lanes D and E). The substitution of a thymine by a guanine in the sLSE of the o(sLSE-RSE) super-operator increased significantly the binding affinity of CopG$_{WT}$ (Table 25). However, this modification failed to promote the binding of protein CopG$_{A30E}$ to the SE (Figure 53B, lanes D and E).

To test whether the spacing between the LSE and RSE subsites was important for binding of CopG$_{A30E}$, the o(sLSE-RSE) operator was extended by the insertion of one T:A base-pair between

both subsites (Figure 52F). This modification did not promote binding of mutant protein $CopG_{A30E}$ to the SE (Figure 53B, lane F); indeed, insertion of one base-pair seemed to impair somewhat the cooperativity of binding of the second dimer of $CopG_{WT}$ to the o(sLSE+1-RSE) operator (Figure 53A, lane F). The alteration of the dimer-dimer interaction surface, and hence the conformation of the nucleoprotein complex corresponding to a tetramer bound to the o(sLSE+1-RSE) operator, is reflected by its slightly higher electrophoretic mobility relative to the o(LSE-RSE) operator carrying the intact SE (Figure 53A, lanes E and F), which could be accounted for by a decreased bending in the super-operator.

It has been shown that CopG exhibits the greatest affinity for the right half of the wild-type operator (*Chapter III,* Table 23), and that mutant $CopG_{A30E}$ is able to bind exclusively to this half. Several right–operator-half variants differing in the sequence and in the spacing between the RSE and RA subsites were tested. $CopG_{WT}$ bound to all the right–operator-half variants (Figure 53A). Relative binding affinities for these variants could be inferred from the comparison of the fraction of complexed DNA in EMSA. The o(RSE-sRA3) super-operator, which harbors two inversely-repeated copies of the 5'-TGCAC-3' sequence, spaced by 6 nt, was the unique operator variant that significantly increased the affinity of binding (Figure 53A, lane J). Binding to the o(RSE-sRA1) operator, which possesses two directly-repeated copies of the 5'-TGCAC-3' sequence spaced by 5 nt, showed a pattern of bound DNA similar to that of the intact right half of the WT operator, the o(RSE-RA) (Figure 53A, lanes G and H). These results were consistent with those obtained for the full-length super-operators, where the change in sRA1 did not affect significantly the affinity of binding of the protein, and that in the sRA3 variant resulted in a 6-fold increase (Table 25). However, inserting or deleting one base-pair (T:A) between the subsites of the o(RSE-sRA1) and o(RSE-sRA3) operator variants, respectively, reduced the binding affinity of $CopG_{WT}$ to lower levels than to the wild-type sequence (Figure 53A, lanes G-K). On the other hand, the pattern of complexes observed with mutant $CopG_{A30E}$ was practically the same as with $CopG_{WT}$, albeit the weaker intensities of bands could be due to the diminished affinity of this mutant for DNA. No band-shifts were observed in lanes I and K of Figure 53B.

Figure 53C shows the competitive EMSAs carried out using each one of these operator variants comprising exclusively two subsites and the full-length WT operator. Clear differences in the electrophoretic mobility of the complexes generated by binding of $CopG_{WT}$ to the two-site operator variants were observed. These differences are likely because of the altered bending of the DNA molecules upon protein binding.

Overall, these results suggest the relevance of the distance between adjacent binding sites as well as of their relative orientation in the process of protein binding to its target DNA.

4. Effect of the presence *in trans* of the super-operators on the pLS1 copy number

The role of CopG on pMV158 plasmid replication is to exert a strict control on the intracellular amount of RepB molecules, thus maintaining the number of plasmid copies within limits that are metabolically favorable for the cell. On this basis, it was interesting to probe the effect of the presence *in trans* of super-operators on the pMV158 plasmid copy number. A super-operator *in trans* was expected to bind CopG molecules with a higher affinity than the wild-type operator carried on pMV158, leading to derepression of the *copG-repB* operon and therefore to an increase of the intracellular number of copies of the pMV158 replicon.

To address this issue, the sLSE and sRA3 super-operators (Figure 48), which showed the highest increment in protein affinity (Table 25), were selected to be cloned into plasmid pC194. See *M&M* (Part II, 9) for details. Figure 54 depicts the recombinant vectors obtained by cloning the 55-bp oligonucleotides comprising the sLSE and sRA3 super-operators into the single *Hin*dIII site of pC194. Recombinant plasmids pC194-sLSE and pC194-sRA3 possess one copy of the corresponding super-operator, whereas pC194-sLSE(2) contains two tandem copies of the sLSE super-operator (Figure 54). These constructions were obtained as described in *M&M,* and then transformed into pneumococcal cells that harbored plasmid pLS1, a pMV158 derivative. Selected transformants were grown and total DNA was extracted from the bacterial cells and loaded onto agarose gels to be visualized (Figure 55).

The results showed that in all cases, irrespective of the super-operator present *in trans*, the copy number of pLS1 in the heteroplasmid strain was about twice as much as the number of copies of the plasmid in the homoplasmid strain (Figure 55 and Table 26). Thus, while pLS1 maintains an *N* = 22 ± 2 in *S. pneumoniae* in the homoplasmid state (del Solar & Espinosa 1992), its average copy number in the presence of the super-operators was of *N* ~40. A direct relation between the affinity of the protein for the super-operators and the copy number of the pMV158 replicon was not observed (Figure 55 and Table 26). In addition, under the same experimental conditions, the presence *in trans* of the WT operator did not significantly increase the number of copies of pMV158 (unpublished data).

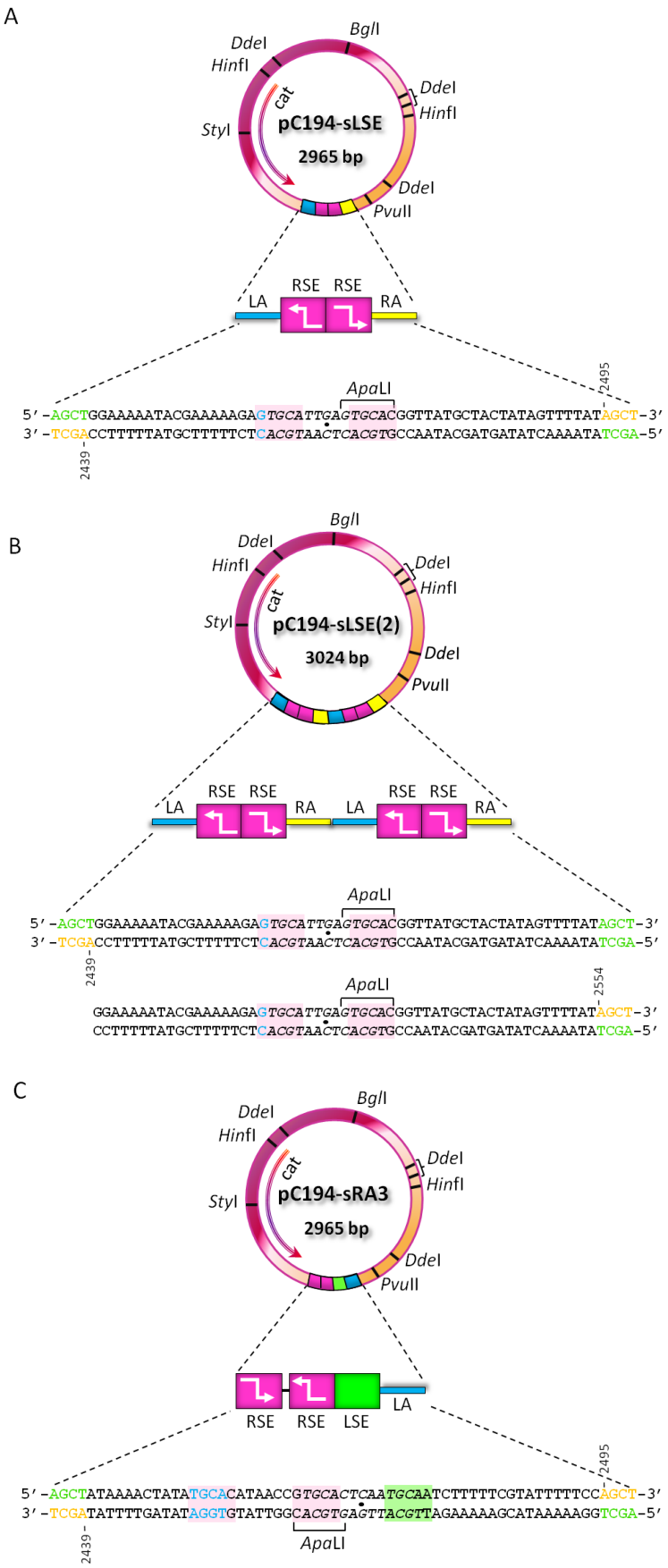


Figure 54. pC194-based recombinant vectors carrying the super-operators sLSE and sRA3. Sequences corresponding to each insert are shown below the pictured vector. Bases recognized by $CopG_{WT}$ in the co-crystals are *green-shaded* in the LSE and *pink-shaded* in the RSE and in the optimized RSE-like binding sites. The 13-bp SE is displayed in *italics* with its two-fold symmetry axis indicated by a *small circle* in all the super-operators. Nucleotides modified in the mutated sequences are in *blue*. At the ends, the *yellow nucleotides* come from the vector and the *green* ones are the protuding nucleotides added to the 55-bp oligonucleotides for cloning in pC194. A*rrows* in the diagrams indicate the orientation of the RSE 5'-TGCAC-3' sequence.

Table 26 | Effect of the presence *in trans* of the super-operators on the copy number of pLS1

Super-operator	Plasmids and copy number[a] in the heteroplasmid state	
in trans	**pC194-based recombinant**	**pLS1**
sLSE	110.3 ± 19.4	39.9 ± 8.3
sLSE(2)	99.1 ± 23.6	41.7 ± 8.4
sRA3	121.2 ± 15.9	40.03 ± 6.7

[a]Copy numbers are the average number of plasmid copies in the different clones analyzed (n = 6).

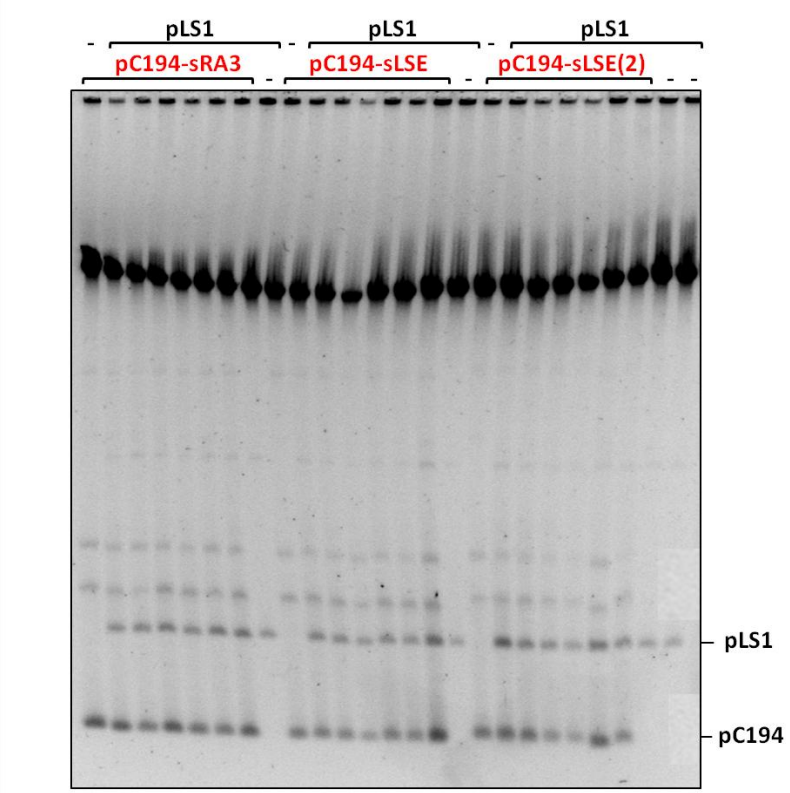


Figure 55. **Total DNA from *S. pneumoniae* bearing pLS1 and the pC194-based recombinants.** The supercoiled monomeric forms of the pMV158- and pC194- plasmid derivatives are indicated.

Studies performed using the super-operators provided us with new data for understanding the specificity of protein binding to its DNA operator. In general, all results suggest that both the nucleotide sequences as well as the distance between the subsites in the operator have an important role in the binding of CopG. On the other hand, the new pC194 derivatives carrying super-operators could be a useful tool for controlling dosage of genes cloned in pMV158-based expression vectors in bacterial systems.

Discussion

The nucleoprotein repression complex at P_{cr} promoter consists of four CopG dimers bound to operator DNA

As mentioned in *Introduction*, the target DNA of CopG spans about 50 bp, overlapping the -35 and -10 hexamers of the P_{cr} promoter (Figure 17). This operator was empirically subdivided into four binding elements: the central LSE and RSE subsites, and the lateral LA and RA subsites (Figures 18 and 56). Studies of analytical ultracentrifugation demonstrated that the whole specific nucleoprotein complex consists in fact of four CopG dimers bound to the target DNA (*Chapter I*) and that the assembly of this complex is highly cooperative because of protein-protein interactions established between adjacent dimers. The 19-bp DNA-protein co-crystals had shown an extensive interaction surface between adjacent dimers bound to each half of the SE (Gomis-Rüth, Solá *et al.* 1998a). All data along this thesis support that each dimer binds to one operator subsite. The characterization of the complex CIV by footprinting assays showed that the four dimers bind to four consecutive regions of the major groove on the same face of the DNA occupying the regions corresponding to these operator subsites (*Chapter III,* Figure 47). Thus, the protection of G589' and G590 against DMS-methylation confirmed the LA as being a site for binding of protein, while the protected bases G619 and G623' (the latter located outside of the proposed interaction sequence) demonstrated the binding of protein to the RA (Figures 47 and 56). The protected bases G600 and G601' implied that the protein binds to the LSE, whereas protection of bases G607 (observed only in the complexes CII of $CopG_{WT}$ and $CopG_{A30E}$ and in the complex CI of $CopG_{G25E}$), G609, G610' and G612' confirmed the binding of protein to the RSE (Figures 47 and 56). These data are consistent with our nucleoprotein model, wherein four CopG dimers bind to the operator, bending 120° the double helix towards the protein-DNA interface (Figure 17).

Additionally, the correlation between the number of complexes generated and the subsites present in the operator DNA strongly suggests that one dimer binds to each subsite. Thus, in experiments performed using mutant operators lacking one of the lateral subsites, i.e. the LA or the RA, the highest-order complex detected was CIII, which would correspond to three CopG dimers bound. On the other hand, when the missing subsite was either of the SE halves, CII was the highest-order complex that accumulated. This fact also highlights the relevance of cooperativity in binding of CopG. Since there was no protein dimer bound to the mutated central site, the cooperative interactions required for binding of the third dimer to the adjacent lateral subsite could not be established and therefore, the assembly of complex CIII was prevented. Alike, in the assays carried out using mutant operators conserving only two adjacent binding sites, CII was the highest-order specific complex that accumulated.

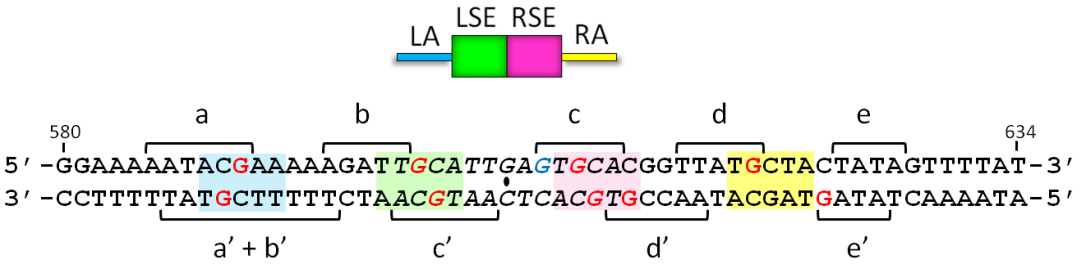


Figure 56. Scheme of the WT operator showing the interactions established by CopG with the DNA backbone and with the G-residues in the complex CIV. Contacted guanines are highlighted in *red*. Guanine in *blue* is protected by $CopG_{WT}$ and $CopG_{A30E}$ in the complexes CII, and by $CopG_{G25E}$ in the complex CI; while it is unprotected in complexes CIII and CIV (Figures 46 and 47). *Brackets* indicate the regions protected by protein against the OH• cleavage (Figures 45 and 47). The self-palindromic sequences 5'-TGCA-3' are in *italics*. The two-fold symmetry axis of the SE is denoted by a *small ellipse*. Coordinates given correspond to the pMV158 sequence. The interaction sequence initially proposed for each operator subsites is *shaded* on the sequence accordingly to the operator diagram (*blue* for LA, *green* for LSE, *magenta* for RSE, and *yellow* for RA).

The assembly of the CopG-operator nucleoprotein complex occurs by sequential and ordered addition of protein dimers

Extensive interaction studies using several operator variants demonstrated that the contribution of individual operator subsites to total affinity of protein for target DNA are different (RSE >> RA ≈ LSE > LA). This data, along with the analysis of cooperative binding to operator variants consisting of two adjacent subsites and the footprinting analysis of the contacts of CopG in the different complexes, allow us to suggest that assembly of the nucleoprotein complex occurs by sequential and ordered addition of dimers to the operator DNA. Thus, CopG would bind first to the RSE, next to the RA, then to the LSE and finally to the LA (Figure 57).

From the estimation of relative affinities, the RSE was shown to be the primary site of CopG binding (*Chapter I*, Table 19,) (Hernández-Arriaga, Rubio-Lepe *et al.* 2009). RSE was the only operator subsite that, when alone, was still able to yield a specific protein-DNA complex. It is worth noting that protein $CopG_{G25E}$ exhibited null cooperative properties, and hence was able to form only complex CI with target DNA. Footprinting analysis of this complex CI demonstrated that, despite the low affinity of $CopG_{G25E}$ for the primary site (about one-10^{th} the affinity of $CopG_{WT}$; Table 22, Figure 36), the RSE remains the unique site on the DNA operator appropriate for the protein specific binding (*Chapter III*). This shows that cooperative contacts are not required for binding of the protein to the RSE. Therefore, the RSE can be considered the first site to be occupied by CopG in the formation of the entire nucleoprotein complex of our working model. This information is important for understanding the transcriptional repression mediated

by CopG through active dissociation of the RNAP-P_{cr} promoter open complexes (Hernández-Arriaga, Rubio-Lepe *et al.* 2009).

Based on the values of the estimated relative affinities (*Chapter I,* Table 19), once the complex CI was formed by binding of CopG to the RSE, a second dimer could bind either to the RA or to the LSE, since both elements seem to contribute similarly to the overall affinity of the protein for the operator. Helpful hints on the second operator subsite to be bound by the protein arose from additional experiments. The analysis of CopG binding to mutant operators carrying only two adjacent subsites showed that $CopG_{WT}$ binds with higher affinity and apparently higher cooperativity to the right half of the operator (RSE and RA subsites) than to the central subsites (LSE and RSE) (*Chapter III*, Table 23, Figures 41 and 42). On the other hand, $CopG_{A30E}$, a mutant protein exhibiting altered cooperative properties and a much lower affinity for the RSE (about one-100th the affinity of the native protein), was unable to form specific complexes of higher order than CII with the entire target DNA. EMSAs performed with mutant operators carrying different combinations of two adjacent subsites demonstrated that $CopG_{A30E}$ binds exclusively to the right half of the operator (Figures 37 and 39A). These results already suggested that RA had a major relevance in protein binding, although only the characterization by footprinting assays of the CII complexes formed either by the wild-type protein or by the $CopG_{A30E}$ mutant demonstrated that the RA is the second site bound by CopG during the sequential binding of the protein to its operator DNA. Therefore, the third subsite to be occupied would be LSE, as confirmed by footprinting analysis of the complex CIII generated by $CopG_{WT}$ (Figure 47).

The last operator subsite to be bound by the protein would be the LA, which contributes the lowest to the global affinity of CopG (*Chapter I,* Table 19). Moreover, footprinting analysis of the complexes generated by the wild-type protein demonstrated that the last site bound to form the complex CIV was this site (*Chapter III*, Figure 47).

So far, our knowledge about the interactions established by $CopG_{WT}$ with the DNA backbone and nucleotides has arisen from the crystallographic studies on the SE (LSE and RSE subsites) bound to two CopG dimers (Gomis-Rüth, Solá *et al.* 1998a), and from the analysis of OH• footprintings of a 179-bp DNA fragment carrying the P_{cr} promoter (del Solar, Pérez-Martín *et al.* 1990). As shown previously, the SE contains two repeats of the self-palindromic sequence 5'-TGCA-3' corresponding to each binding subsite (Figure 57A). Although each CopG dimer has local two-fold symmetry, the two β-strands contact different bases within each SE subsite, not matching the self-symmetry (5'-TGCA-3') of the DNA sequence (Figure 16) (Gomis-Rüth, Solá *et al.* 1998a).

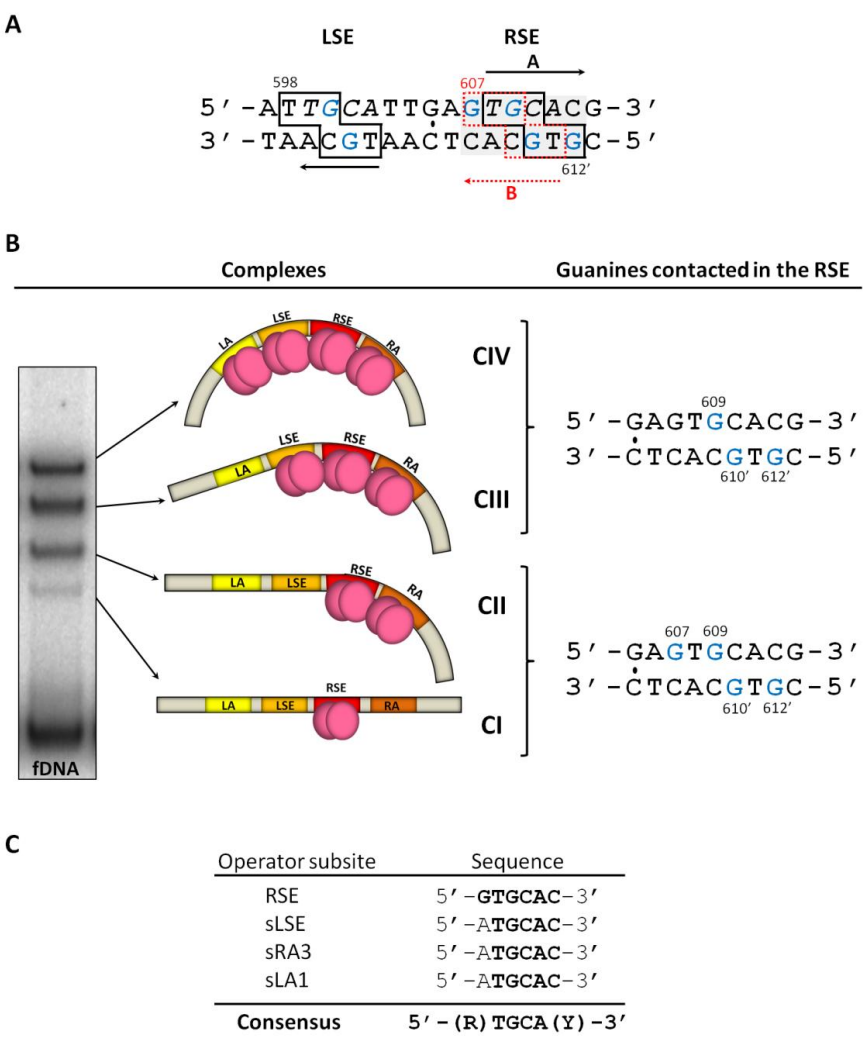


Figure 57. Bases specifically contacted by CopG in the co-crystal with the SE, and the RSE-guanines contacted by CopG in different complexes with the WT operator. (A) Nucleotide sequence of the SE showing the bases contacted by CopG in the co-crystal with the 19-bp oligonucleotide (*filled line boxes*) (Gomis-Rüth, Solá *et al.* 1998a). Guanines are highlighted in *blue*. Because the self-palindromic sequence in the RSE extends 6 bp: 5'-GTGCAC-3' (*grey-shaded*), two possible identical sequences 5'- TGCAC-3' for specific interaction with the protein are found. Namely, the position and orientation indicated by the arrow "A", which corresponds to that observed in the co-crystal (sequence *boxed with black filled line*, wherein the G609, G610 'and G612' would be contacted), and the position and orientation indicated by arrow "B" (sequence *boxed with red dashed line*, wherein the G607 would be contacted while the G612' would not). *Arrows* indicate the direction of the sequence 5'-TGCAC-3'. The self-palindromic sequences 5'-TGCA-3' are in *italics*. The two-fold symmetry axis of the SE is denoted by a *small ellipse*. (B) Scheme showing the guanines (*blue*) observed by footprinting assays to interact with the protein in the different complexes. Coordinates given correspond to the pMV158 sequence. The four operator subsites are colored depending on their contribution to the global affinity of the protein (*red*, the most; *yellow,* the least). (C) Sequences of the RSE in the WT operator, and that of the improved subsites in the sLSE, sRA3 and sLA1 super-operators. *Bold* bases denote identity with the RSE-sequence. In the consensus sequence, *R* and *Y* indicate a purine and a pyrimidine respectively.

It has been reported that the bulk of the sequence specificity comes from hydrogen bonding interactions between the protein and the DNA. Bidentate interactions, in which a single amino acid side chain forms two hydrogen bonds with the DNA, confer an even higher degree of

specificity than single hydrogen bonds to a side chain. For example, arginine can recognize a guanine base through bidentate interactions. There is no other DNA base that contains two hydrogen bond acceptors in the major groove, therefore no base substitutions can be made at this position without reducing DNA binding affinity. Indeed, arginine-guanine pairings are common in protein-DNA complexes, as are lysine-guanine (Luscombe, Laskowski *et al.* 2001).

The analysis of the nucleotide sequence in the RSE allows us to hypothesize the molecular basis that may render this region the primary binding site of CopG. First, the nucleotide sequence of RSE has an extended internal symmetry, which gives an *Apa*LI restriction site (5'-GTGCAC-3') (Figure 57A). Therefore, and based on the specific contacts made by protein in the 19-bp dsDNA co-crystals, a CopG dimer could recognize two alternative although identical overlapping 5'-TGCAC-3' sequences within the RSE: 1) that in position and orientation "A", as reported in the crystallographic studies (Gomis-Rüth, Solá *et al.* 1998a), and 2) that in position and orientation "B", 1 nt closer to the SE two-fold symmetry axis (Figure 57A). In both cases, the number of guanines contacted by the protein would exceed that in the LSE. CopG could establish bidentate interactions through Arg4 (from either monomer) with three guanines of either of these alternative sequences in the RSE, while only with two guanines in the LSE (Figure 57A). Consistently, when the thymine 598 was substituted by a guanine in the sLSE super-operator, a great increase (about 18-fold) in protein affinity for the operator was observed (*Chapter VI*, Table 25). This indicates the relevance of the strong interactions made between the arginine residues and guanines in the specific recognition of the RSE.

The protection patterns observed by footprinting would support the hypothesis of the possible recognition of two binding positions in the primary site (*Chapter III*, Figure 47). However, preliminary results from NMR assays suggest that a CopG dimer does not bind mostly to either of these options (A or B), but seems to recognize the entire sequence 5'-GTGCAC-3' that includes the two alternative sets of the above-mentioned specific interactions. This evidence indicates that a protein dimer could establish an optimal interaction of high affinity and specificity through the recognition of the 6-bp sequence 5'-GTGCAC-3' in the RSE. The recognition of this extended motif might involve that the CopG dimer, having local two-fold symmetry, establishes symmetric contacts with the operator DNA, as it occurs with MetJ, where the interactions made by the side chains of Lys23 and Thr25 from each monomer are totally symmetric related to the metbox sequence (Somers & Phillips 1992) (Figure 6A). In the case of CopG, the side chains of two Arg4 (from each monomer) could make strong contacts with the four guanines in the RSE, thus positioning the dyad axis of the protein dimer with that of the DNA sequence. This class of

interaction, wherein a protein dimer specifically recognizes a self-palindromic sequence on DNA, is not a common feature among the RHH proteins. This would also be compatible with the DMS footprinting results of the complex CI formed by $CopG_{G25E}$ (*Chapter III*, Figure 47D). In addition, this extended sequence would remain contacted upon binding of the second dimer to the RA.

The high-resolution footprintings revealed that in the complexes CII, formed by a tetramer bound to the right half of the operator, the four guanines of the 5'-GTGCAC-3' sequence were contacted either by $CopG_{WT}$ or by $CopG_{A30E}$ in the RSE (*Chapter III*, Figure 47). However, in the co-crystals formed by a CopG tetramer bound to the SE, interactions with the guanine at position 607 (G607) were not observed (Figure 57A) (Gomis-Rüth, Solá *et al.* 1998a). Moreover, the contact with G607 disappeared once the complex CIII was formed and was also missing in the complex CIV (Figure 57B). This suggests that the binding of another dimer to the LSE constrains the position of the protein in the RSE, as if only the "position A" allows the appropriate dimer-dimer interactions in the tetramer bound to the SE.

On the other hand, studies using the super-operators gave us useful information supporting the hypothesis of the 6-bp recognition sequence in the primary site. In this work, the RSE 5'-TGCAC-3' sequence, which was contacted by CopG in the co-crystals of the tetramer bound to the 19-bp SE (Gomis-Rüth, Solá *et al.* 1998a) was taken as a reference to make the base substitutions and alter the orientation of the sequence in the 'improved' subsites of the super-operators (*Chapter IV*, Figure 48). However, the relative binding affinities observed for these operator variants, as well as the analysis of their improved sequences, also suggest that the specific sequence recognized by the protein might extend further than that we have considered for creating the RSE-like sites. Hence, taking into account the newly-assumed recognition sequence for CopG in the primary site (5'-GTGCAC-3'), we observed similarities in the improved sites of the sLSE, sRA3 and sLA1 super-operators, whose "modifications" in the sequence showed evident increases in the affinity of protein for the target DNA (Figures 48 and 57C, Table 25). Thereby, we propose the consensus sequence 5'-(R)TGCA(Y)-3' (R and Y for unspecified purine and pyrimidine, respectively) for promoting better the binding of CopG to DNA. Accordingly, the super-operators harboring 'improved' sequences that do not match with the consensus 5'-(R)TGCA(Y)-5' showed a null effect in the affinity of protein (Figure 48, Table 25).

All together, these results allow us to conclude that the high affinity of protein for the RSE relies mainly on the nucleotide sequence. The RSE is the only site in the operator that offers a sequence highly suitable to form strong hydrogen bonds with the residues of the protein. On the other hand, we reasoned that CopG would bind secondly, and with high cooperativity, to the RA

because the resulting configuration of the tetramer bound does not disrupt the whole strong interactions made firstly at the RSE. However, when the third dimer binds to the LSE, the contacts in the RSE rearrange to become those reported from the co-crystals of a tetramer bound to the SE (Figure 57B) (Gomis-Rüth, Solá *et al.* 1998a). This hypothesis is supported by preliminary data of NMR, which show that the binding of the second dimer to the RA, either in the case of $CopG_{WT}$ or $CopG_{A30E}$, does not provoke changes in the interactions already established by the dimer bound to the RSE (not shown). We could also suggest that the interactions of the protein with the nucleotides that we report in this work are more likely to occur *in vivo* than the interactions arising from the co-crystallization studies, where the complex CII was "forced" to be integrated by two dimers bound to the LSE and RSE subsites. Furthermore, it would be very interesting to obtain the co-crystals of a CopG dimer bound to the separate RSE.

There might be at least two different interdimeric surfaces in the nucleoprotein complex

The characterization of the complexes between the CopG variants and a number of target DNAs corroborates that, besides the specific recognition of individual binding sites, the establishment of appropriate dimer-dimer interactions plays a key role in binding of this repressor. These data, along with the relative affinities observed both for the individual subsites (*Chapter I*, Figure 25, Table 19) and for pairs of adjacent operator subsites (*Chapter III,* Figure 40, Table 23), suggest that a CopG tetramer is formed on each half of the operator DNA upon binding of the protein.

The dimer-dimer interaction surface in the tetramer bound to the right half of the target DNA seems to differ from that formed at the center of the operator between both tetramers, as suggested mainly by: i) the higher apparent cooperativity of binding of a second CopG dimer to the RA than to the LSE (Figure 41, Table 24); ii) the conformational differences between nucleoprotein complexes carrying a CopG tetramer at the right operator half and at the SE, which are reflected by a significantly lower relative mobility of the complex at the SE (Figure 53); and iii) the ability of $CopG_{A30E}$ to form a tetramer at the RSE-RA tandem sites but not at the LSE-RSE operator DNA (Figures 37, 39A and 53B, Table 23). Simulation of the conformation of a $CopG_{A30E}$ tetramer bound to the SE, based on the co-crystals and preliminary NMR analysis of the $CopG_{WT}$ protein, gives us clues on the possible molecular basis for the inability of the mutant protein to bind to both central subsites (Figure 58). Substitution of Ala30 by Glu in the second helix of $CopG_{A30E}$ generates an electrostatic repulsion between the negatively-charged side chains of the Glu-residues and the DNA phosphodiester backbone that would prevent the proper

binding of $CopG_{A30E}$ to the SE (Figure 58). The fact that the mutant protein does form a tetramer upon binding to the right operator half strongly suggests a different arrangement of the protein dimers relative to each other and to the DNA at this region, so that the electrostatic repulsion is prevented in the $CopG_{A30E}$ nucleoprotein complex. We do not know yet whether this effect may result from a different bending or kinking of the DNA molecule upon binding of the protein to either the SE or the right operator half, as suggested by the differences in electrophoretic mobility of the corresponding complexes (Figure 53).

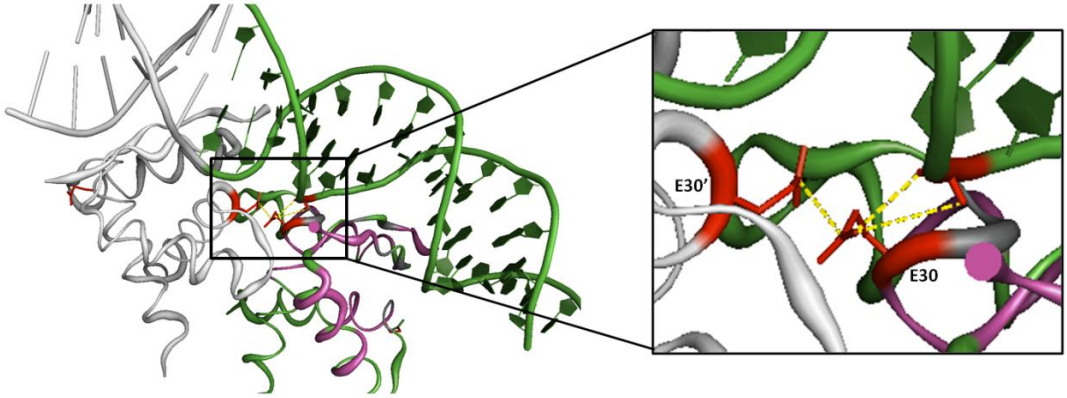


Figure 58. Simulation of the $CopG_{A30E}$ tetramer bound to the SE. Based on data obtained for the wild-type protein bound to an 18-bp oligonucleotide (ribbon plot of DNA in *green*) carrying the RSE alone (preliminary RMN data) and on the interactions made by a $CopG_{WT}$ tetramer bound to the SE in the co-crystals (Gomis-Rüth, Solá *et al.* 1998a), a tetramer of the mutant protein bound to the SE was simulated. The c*lose-up view* shows the repulsive interactions (*yellow sticks*) between the negatively charged Glu30 residues (E30, *in red*) of two $CopG_{A30E}$ dimers and the DNA phosphodiester backbone of the SE.

Less data are available on the dimer-dimer interaction surface of the CopG tetramer bound to the left operator half, since the lack of the primary site precludes binding of $CopG_{A30E}$ and makes binding of $CopG_{WT}$ depend on the nearly simultaneous binding of two dimers, hence giving an appearance of huge cooperativity (*Chapter III,* Figures 41 and 42). Several pieces of information point, however, to the tetramer conformation at the left operator half being similar to that at the right half.

By EMSA using the operator variant lacking the primary site (RSE^-) and different concentrations of $CopG_{WT}$ (*Chapter I*, Figure 25, Table 19), the appearance of complex CII indicates that a tetramer should be bound to the left half of the operator. Since missing of the RSE prevents the protein from making cooperative contacts that would promote binding of a further dimer to the RA, complex CIII could not be formed. In contrast, complex CIII was observed when the missing subsite was either of the lateral ones (RA^- and LA^-) (Figure 25). However, if the protein

concentration was high enough, complex CIV was generated with high cooperativity on the RSE$^-$ operator, as though the two remaining dimers bound simultaneously to DNA, thus allowing the tetramers bound to each operator half to interact through the central dimers, one of which has become bound despite the absence of the RSE (not shown). On the other hand, improving the LA by turning it into a RSE-like site (the o(sLA1-LSE) super-operator) increased the binding affinity of the $CopG_{WT}$ and enabled binding of $CopG_{A30E}$. However, $CopG_{A30E}$ was unable form a tetramer at the SE even though the nucleotide sequence of the LSE was improved (the o(sLSE-RSE) super-operator) or the spacing between subsites was enlarged in 1 nt (the o(sLSE+1-RSE) variant) (*Chapter IV*, Figure 53). This indicates that the deficiency of mutant $CopG_{A30E}$ on binding to the SE is not due to the decrease in sequence specificity but rather to unfeasible dimer-dimer interactions. This fact was a strong indication of the existence of an inter-dimeric surface different from that observed in the co-crystals, within the entire nucleoprotein complex of CopG with its target DNA. Besides, according to their electrophoretic mobility, the complexes formed by $CopG_{WT}$ on the two-subsite super-operators of the left operator half (the o(sLA1-LSE) and o(sLA1-1-LSE) variants) seemed to have a conformation similar to that corresponding to a protein tetramer bound to the right half of the target DNA (*Chapter IV*, Figure 53). Clearly, binding experiments employing a two-site operator variant that carries the LA and the improved LSE are missing. However, the results obtained with the o(sLA1-LSE) super-operator are interesting and support the hypothesis of the formation on each operator half of tetramers that would have an interdimeric surface different from the one established between the dimers bound to the SE. The o(sLA1-LSE) operator, which is able to accommodate a tetramer of $CopG_{A30E}$, has a RSE-like site adjacent to the LSE. Thus, the mutant protein cannot bind to the LSE in the SE but it does in the o(sLA1-LSE) (Figure 53B, lane B). Hence, although possessing a pair of sites equivalent to those in the SE, the o(sLA1-LSE) variant shows a different sequence and size of the inter-subsite spacer (see below) that does allow the binding of $CopG_{A30E}$, demonstrating the relevance of making different dimer-dimer interactions. On this basis, we would have expected this protein to be able to generate the complex CIV on the full-length sLA1 super-operator. However, despite sLA1 carrying the improved sequence at the left half and conserving the wild-type right half (both separate operator halves being able to accommodate a $CopG_{A30E}$ tetramer), the electrostatic hindrance of this mutant protein in binding to the SE prevented the formation of complex CIV. This reinforces the hypothesis that there could be at least two different dimer-dimer interfaces in the whole nucleoprotein complex generated by the wild-type protein, namely, the inter-tetrameric surface, equivalent to the dimer-dimer interface observed

in the co-crystals of the protein bound to the SE (Gomis-Rüth, Solá *et al.* 1998a), and the intra-tetrameric surface between the two dimers bound to either of the operator halves (Figure 59).

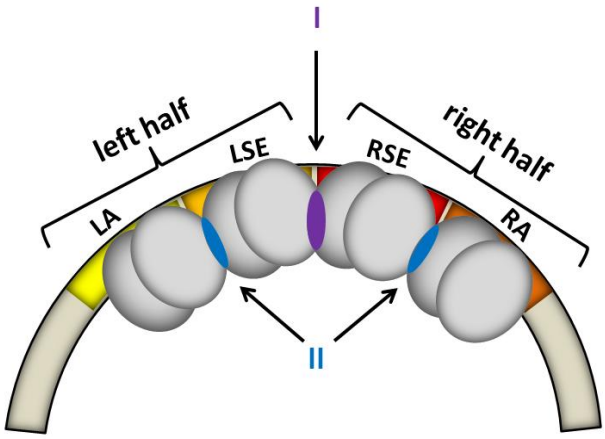


Figure 59. The two putative interdimeric surfaces generated in the nucleoprotein complex. Picture of the whole CopG-DNA operator nucleoprotein complex showing the two putative interfaces: I, the inter-tetrameric surface (*violet*) and II, the intra-tetrameric surface (*blue*). The four operator subsites are colored depending on their contribution to the global affinity of protein (*red*, the most; *yellow,* the least).

Binding dependence on the nucleotide sequence and on the dimer-dimer contacts established upon binding to the operator sites is a well-known characteristic of the RHH regulators. The effect of altering the spacing or orientation of the subsites in the binding affinity has been well studied in some RHH regulators, such as Arc (Smith & Sauer 1995), the *E. coli* F-factor TraY protein (Lum, Rodgers *et al.* 2002; Williams & Schildbach 2007), the antitoxin RelB (Overgaard, Borch *et al.* 2009) and the regulator ω (de la Hoz, Pratto *et al.* 2004; Weihofen, Cicek *et al.* 2006; Dostál, Pratto *et al.* 2007). Mostly, spacing and orientation other than the wild-type eliminates the cooperative interactions and the binding affinities of RHH proteins for their operators. For instance, the ω repressor binds cooperatively to promoters with seven to ten consecutive heptad repeats (5'-(A/T)ATCAC(A/T)-3') in palindromic inverted, converging or diverging orientation, and also to promoters in non-palindromic direct repeats (de la Hoz, Pratto *et al.* 2004). However, different numbers and orientations of the heptads lead to different distances between the α1-helices of adjacent ω dimers, thereby modulating cooperative interactions between the protein and the different operators (Weihofen, Cicek *et al.* 2006).

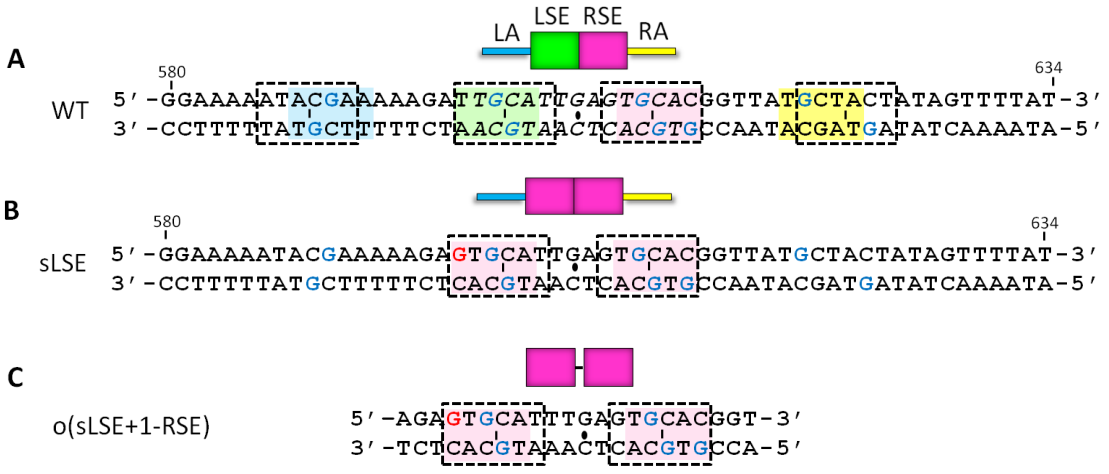
In the case of CopG, spacing between the subsites seemed also to affect somehow the affinity of protein for the operator DNA (Figure 53), albeit more studies should be addressed to make conclusions. It is important to remind that, as stated in *Chapter CIV*, the orientation of subsite sequence in the super-operators made reference to the motif 5'-TGCAC-3', which was employed

to optimize the operator subsites. However, the analysis of the data obtained in this work points to the 6-bp consensus sequence 5'-(R)TGCA(Y)-3' being actually more relevant in the binding affinity of the protein (see above). Since this is a self-palindromic sequence, the orientation of the subsites would no longer make sense, and the sequence and size of the inter-subsite spacing should be considered instead. In Figure 60 the super-operators that provide some information about the effect of the inter-subsite spacing on CopG binding are analyzed. In addition, a detailed inspection of the sRA3 and sLA1 super-operators, which were improved in the right- or left- half of the operator, respectively, suggests that the recognition sequences in the lateral subsites (LA and RA) of the operator could be other than those firstly taken into account (Figure 56 and 60A).

The poor matching with the consensus 5'-(R)TGCA(Y)-3' in the lateral subsites confirms that specificity of sequence is not entirely responsible for binding of CopG to the target DNA (Figure 60A). Thus, in the case of the left operator half, the sequence that best matches with the consensus within the LA, at least with respect to the chemical nature of the nucleotides, is 5'-ATACGA-3'. This 6-bp motif would include the two guanines (G589' and G590) contacted by the protein as observed in the footprintings and would be spaced center-to-center from the LSE by 12 nt (Figure 60A). The changes made in that sequence in order to improve the LA in the sLA1 super-operator increased 4-fold the binding affinity of CopG for its target DNA, which could indicate that these 6-bp (5'-ATACGA-3') were in fact the putative recognition sequence for the protein in the LA (Figure 60A and D). Another evidence is that the deletion of one base-pair (turning the spacing to only 11 nt) between the improved LA and the LSE (the o(sLA1-1-LSE) operator) caused the loss of the increment in binding affinity observed for the o(sLA1-LSE) super-operator (Figure 53A), as though the proximity of these subsites, despite of having a good sequence in the LA, hinder the proper binding of the CopG tetramer to the left operator half (Figure 60E). This could also explain the insignificant effect on protein binding observed for the sLA2 super-operator (Table 25), in which the best matching with consensus in the LA (5'-GTGCAA-3') would be also separated from the LSE by 11 nt (Figure 60F).This hypothesis is consistent with the binding of the mutant $CopG_{A30E}$ to the o(sLA1-LSE) but not to o(sLA1-1-LSE) (Figure 53B).

In the case of the right operator half, the unique super-operator that increased the affinity of binding by the protein (a 6-fold increase) was sRA3 (Table 25). In this operator variant, the new RSE-like site (5'-GTGCAT-3') is separated from the RSE by a center-to-center spacing of 12 nt (Figure 60G). This fact, together with the data for the left operator half, suggests that an inter-

subsite spacing of 12 nt induces a configuration of the tetramers bound to either of the operator halves that favors somehow the affinity of protein (Figure 60). Thereby, following the argument employed for the LA subsite, the sequence 5'-GCTACT-3' could be the recognition site for CopG in the RA. This motif would include the two guanines (G619 and G623') contacted by the protein, as observed by footprintings, and would be separated from the RSE by 12 nt, a center-to-center spacing that seems to be appropriate for binding of tetramers to the operator halves (Figure 60). Moreover, the deletion of one base-pair (turning the spacing to 11 nt) between the RSE and the improved RA (the o(RSE-1-sRA3) variant) reduced the affinity of binding to the protein to lower levels than the affinity observed for the wild-type right operator site (Figure 53A). This fact indicates that, irrespective of the 'good' sequence, the shorter inter-subsite spacing prevents the proper binding of the CopG tetramer to the right half (Figure 60H). Regarding the sRA1 super-operator, which was not significantly affected in the affinity of binding to the protein (Table 25), has RSE and the 'improved' RA (5'-TGCACC-3') separated by a center-to-center spacing of 11 nt (Figure 60I); however, the insertion of one base-pair (turning the spacing to 12 nt) between both subsites (the o(RSE+1-sRA1) variant) reduced the affinity of binding by $CopG_{WT}$ (Figures 53A and 60J). On the other hand, in the case of the sRA2 and sRA4 super-operators, which neither affected the affinity of binding (Table 25), a key role of the nucleotide sequence seems to be involved. Thus, in sRA2, the substitution of a T:A base-pair by a C:G one did not increase the affinity of protein binding compared to the wild-type operator, despite of adding a guanine in the region of RA (Figure 60K). Meanwhile, swapping the C and A bases of the 5'-GTGCAT-3' sequence at the RA region of the efficient super-operator sRA3 slightly decreased the affinity of binding to the protein (sRA4 super-operator) (Table 25, Figure 60L). These results also indicate the relevance of the sequence of bases in the subsite that is specifically contacted by CopG.



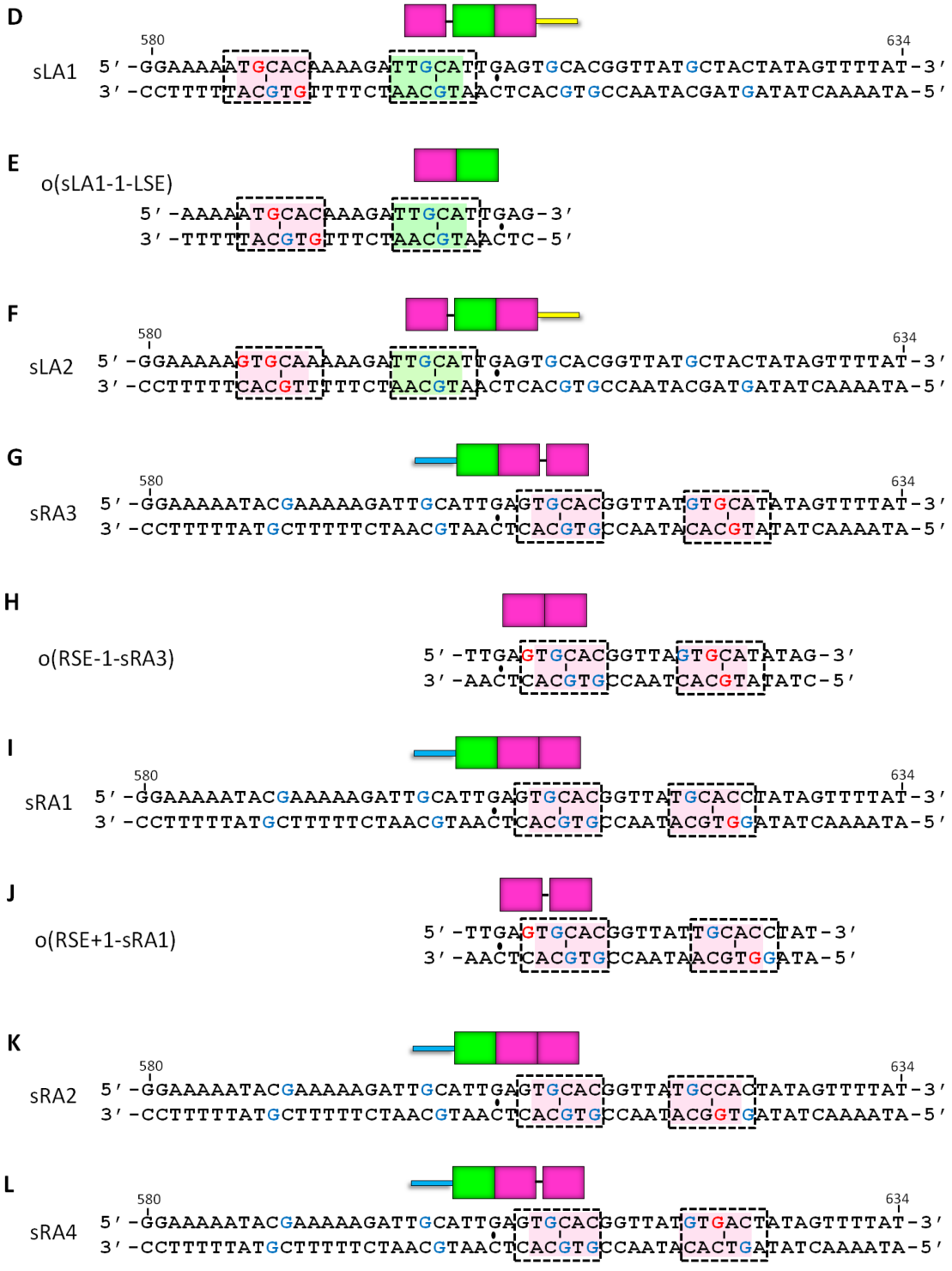


Figure 60. Analysis of the sequence and inter-subsite spacing in the WT operator and different super-operators. The 13-bp SE is displayed in *italics* in the WT operator (A) with its two-fold symmetry axis indicated by a *small ellipse*. The initially proposed operator subsites are shaded on the sequence accordingly to the operator diagrams: *blue* for LA, *green* for LSE, *magenta* for RSE and for the RSE-like sites in the super-operators and *yellow* for RA. The new putative recognition sequences based on the consensus 5'-(R)TGCA(Y)-3' (R and Y for unspecified purine and pyrimidine, respectively) are *boxed with dashed lines* and a *small line* indicates their two-fold symmetry axes. The guanines observed to be protected against DMS methylation by $CopG_{WT}$ in complex CIV are in *blue.* The guanines of the super-operators that are likely to contact with protein are in *red*. Coordinates correspond to the sequence of pMV158.

In the case of the SE subsites, as shown, the self-palindromic 5'-GTGCAC-3' would be the entire recognition sequence of the protein in the primary site (see above), while in the LSE, the best matching with the consensus is 5'-TTGCAT-3' (Figure 60A). These two subsites would be separated by a center-to-center spacing of 9 nt, the same as in the sLSE super-operator (Figure 60B). This closer arrangement of the central subsites of the operator might both favor the formation of the inter-tetrameric interface with the wild-type protein and prevent the formation of a tetramer by the $CopG_{A30E}$ mutant. Thus, when the spacing between the RSE and LSE subsites was lengthened one base-pair (the o(sLSE+1-RSE) operator), the affinity and cooperativity of binding to the wild-type protein decreased, and neither allowed the binding to the mutant $CopG_{A30E}$ (Figure 53 and 60C). The nucleotide sequence of the inter-subsite spacers, because of its inherent bending and flexibility, may also have a major role in defining the dimer-dimer interface, and hence the cooperative interactions in the binding of the protein. This point should be addressed in the near future in order to gain more information on the putative different inter-dimeric surfaces formed upon binding of CopG to its operator.

Summarizing, all these results together suggest that: i) the configuration acquired by the tetramer bound to either of the operator halves, and hence the interactions made between the dimers, is affected both by the sequence of bases at the specific subsites, as well as the length and nucleotide sequence of the intervening DNA; ii) the interdimeric surface created between the dimers at the operator halves should be quite similar, if not equivalent, as seen by comparable features observed in the sRA3 and sLA1 super-operators, and iii) a shorter distance (of only 9 nt) between the SE subsites supports an interaction surface different from that generated between the dimers forming the tetramers at each of the operator halves (Figure 59).

Implication of the CopG sequential binding in the transcriptional repression mechanism at the P_{cr} promoter

Inhibition of transcription initiation in bacteria can take place at any of the steps in the process leading to the formation of the elongation complex (Rojo 1999; Rojo 2001). Although the best-characterized transcriptional repressors seem to act by hindering the access of the RNAP to the promoter, an increasing number of regulatory proteins have been reported that inhibit transcription initiation at a step subsequent to the RNAP binding. In our case, CopG represses the initiation of transcription from P_{cr} by acting at two different steps in the process of formation of an open complex (Hernández-Arriaga, Rubio-Lepe *et al.* 2009). Some data reported here contributed to understand one of the mechanisms of transcriptional repression mediated by CopG. First, CopG competes with the RNAP for binding to the same region of the DNA molecule,

where promoter and operator overlap. At this stage, CopG bound to its target DNA would hinder the access of the RNAP to the promoter region by a steric exclusion mechanism, which constitutes a classic model of repressor action. Second, CopG is able to displace the RNAP stably bound to the P_{cr} promoter. A model for the mechanism of active dissociation of the stable RNAP-P_{cr} complexes was proposed (Figure 61) (Hernández-Arriaga, Rubio-Lepe *et al.* 2009) and several features of CopG repressor that were studied here support it. For this repression mechanism, the highly specific binding to the RSE and the sequential and highly cooperative binding of CopG dimers to the four subsites constituting the operator, hence generating a stable nucleoprotein P_{cr} promoter repressosome through dimer-dimer interactions, seem to be critical.

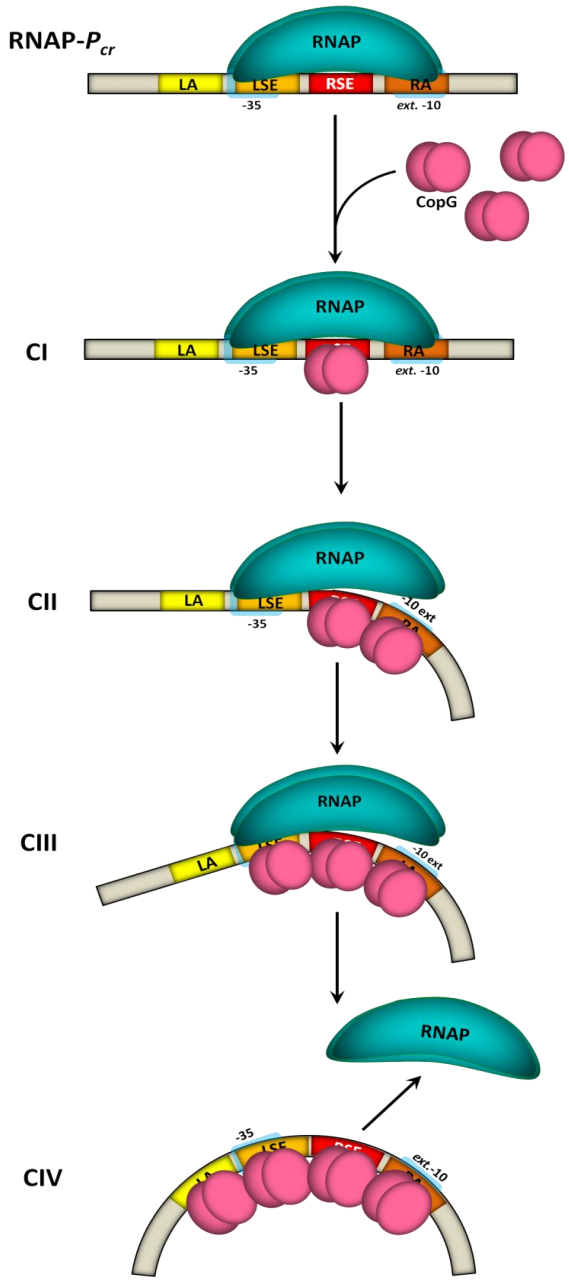


Figure 61. Model for RNAP displacement from P_{cr} mediated by CopG binding to its target DNA. Scheme illustrating one of the two mechanisms of inhibition of transcription mediated by CopG. CopG binds with high affinity to the RSE, which remains accessible in the RNAP-P_{cr} complexes. This association promotes the consecutive and rapid binding of other dimers, until complex CIV is formed. Distortion of DNA (bending of ~120°) induced by repressor binding could help to destabilize the RNAP provoking its release from the promoter. The four operator subsites are colored depending on their contribution to the global affinity of protein (*red*, the most; *yellow,* the least). The -35 and *ext.* -10 regions of the promoter are indicated by transparent *blue boxes.*

On the one hand, we have shown that the primary site of CopG binding is the RSE, to which the protein binds firstly in the sequential addition of CopG molecules to the operator DNA. Moreover, the protein establishes more specific interactions with the RSE than with any other subsite in the entire complex (Figures 47 and 56). On the other hand, this subsite remains accessible to the repressor even in the stable RNAP-P_{cr} complexes since at the RSE region there are only two phosphate groups contacted by both proteins and there are no expected contacts of RNAP with specific bases of the DNA (Figure 62). This situation is unique for the primary site of CopG, because the adjacent LSE and RA subsites, where subsequent repressor molecules would bind, contain respectively, the -35 and the -10 hexamers of the P_{cr} promoter, as well as a higher number of phosphate groups contacted by CopG and the RNAP (Figure 62) (Hernández-Arriaga, Rubio-Lepe *et al.* 2009).

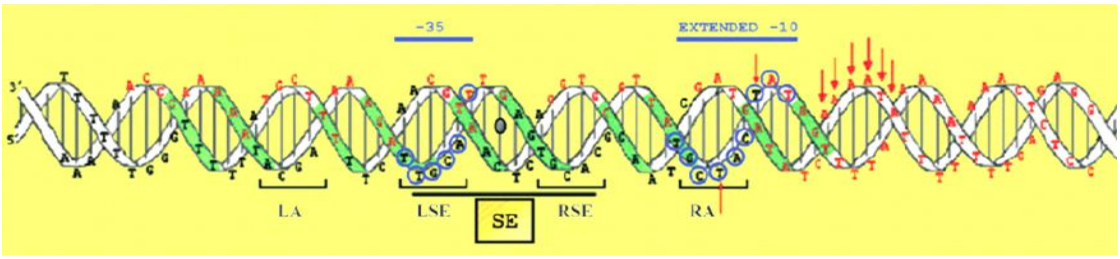


Figure 62. Schematic representation of the contacts of CopG and RNAP with the DNA backbone of the P_{cr} promoter region at 37 °C. RNAP contacts are denoted by *red letters*. *Red arrows* indicate positions in the RNAP-DNA complex with enhanced backbone cleavage. Contacts by CopG are *green-shaded*. The SE is *boxed.* The binding subsites of CopG dimers are indicated with *brackets*. The -35 and *ext.* -10 sequences of P_{cr} are denoted by *blue encircled* letters. Taken from (Hernández-Arriaga, Rubio-Lepe *et al.* 2009).

Thus, the suggested sequence of events in the model for CopG-induced disassembly of the stable RNAP-P_{cr} complexes, which also takes into account the small size of CopG (45 amino acids per subunit), would be as follows (Figure 61): a repressor dimer binds to the RSE, the primary site of binding for CopG, establishing contacts with specific bases of the DNA through the major groove. This preliminary CopG-operator interaction might affect the conformation of the DNA, thus weakening the contacts between the RNAP and the nucleic acid. If repressor concentration is high enough, and because of the highly cooperative properties of CopG, the binding of the first dimer to the RSE would promote the sequential and ordered binding of dimers to the rest of subsites in the operator until assembling the entire nucleoprotein complex. A second dimer would bind to the RA subsite, blocking the extended -10 of P_{cr} promoter, and without altering the specific contacts made firstly at the RSE. Then, a third CopG molecule would bind to the LSE subsite. The addition of this dimer is likely to provoke changes in the conformation of the RSE-

bound dimer so that some contacts with the nucleotide bases disappear. However, the binding of the protein to the LSE would promote the binding of the fourth dimer to the LA subsite. The four dimers would arrange as two interacting tetramers, each bound to a half of the operator, inducing a strong bend of the nucleic acid toward the repressor-DNA interface, which could destabilize binding of the RNAP until finally displace it from the promoter.

The dual repression mechanism mediated by CopG seems to be appropriate for a control system involved in regulation of plasmid replication, as variations in the intracellular conditions would be rapidly sensed due to the short half-life of the CopG/operator complexes (~14 s) and to the ability of CopG to dislodge the bound RNAP (Hernández-Arriaga, Rubio-Lepe *et al.* 2009). The high specificity of the RSE, which is the unique site along the plasmid DNA that offers a proper sequence to be specifically contacted by a two-fold symmetry protein (constituted by no more than the RHH motif), as well as the highly cooperative properties of CopG in the binding to its operator DNA, underlie the effectiveness of this repressor in the tight control of pMV158 plasmid replication.

Concluding remarks

From this experimental work we draw the following CONCLUDING REMARKS:

1. The pMV158 transcriptional repressor CopG binds cooperatively to the operator DNA, and generates a specific nucleoprotein complex consisting of four dimers bound to four operator subsites in the target DNA. These binding sites are the half-sites of the SE (LSE and RSE), and two lateral elements (LA and RA) on either side of the SE.

2. The formation of the nucleoprotein complex occurs by a sequential and ordered binding of CopG dimers to the DNA operator subsites. Thus, the protein binds firstly to the RSE, which was determined to be the primary site for protein binding; then to the RA, next to the LSE and finally to the LA.

3. The mutants $CopG_{G25E}$ and $CopG_{A30E}$ are impaired in their affinity and cooperative properties. Thus, while $CopG_{WT}$ generates four well-defined complexes when bound to its target DNA, $CopG_{G25E}$ forms only the complex CI (a dimer bound to the RSE) and $CopG_{A30E}$ generates a complex CII corresponding to a tetramer bound to the right half of the operator DNA.

4. The unsuccessful binding of mutant $CopG_{A30E}$ to the central subsites in the operator is most likely due to an inappropriate dimer-dimer interaction surface rather than to a lower binding affinity to the sequence of the LSE relative to that of the RA. An electrostatic repulsion created between the negatively charge side chains of the Glu-residues at position 30 and the DNA backbone appears to be involved on the failed binding of this mutant to the SE.

5. The solution three-dimensional structure of wild-type CopG was quite similar to that reported from the co-crystallization studies. Moreover, the two mutant CopG proteins were shown to exhibit identical elements of secondary structure as compared to the native protein. The three-dimensional configuration of mutant $CopG_{A30E}$ was fully solved and fits with that of the wild-type protein.

6. The thermal stability of the three CopG variants depends on the concentration but varies as function of the protein. Thus, mutant $CopG_{A30E}$ is more stable than $CopG_{WT}$ against the increment of temperature, and this latter, in turns, is more resistant to thermal denaturation than mutant $CopG_{G25E}$.

7. The binding affinity of the protein to the operator DNA is sequence-specific. The whole recognition sequence in the RSE might correspond to the self-palindromic 6-bp motif 5'-GTGCAC-3', and we propose the sequence 5`-(R)TGCA(Y)-3' as the consensus for the high affinity binding of CopG.

8. We also proposed that, upon binding, the four CopG dimers arrange as two interacting tetramers, each bound to a half of the operator DNA. This arrangement would generate two different dimer-dimer surfaces in the nucleoprotein complex: one, the inter-tetrameric surface, i.e. between the LSE- and RSE-bound dimers, and the second, the intra-tetrameric surface, i.e. between the dimers forming the tetramers bound to each operator half.

De este trabajo experimental se desprenden las siguientes CONCLUSIONES:

1. El represor transcripcional CopG se une cooperativamente a su DNA operador y genera un complejo nucleoproteico específico que consta de cuatro dímeros de proteína unidos a cuatro subsitios del DNA diana. Los sitios de unión de CopG son: las dos mitades del Elemento Simétrico (subsitios LSE y RSE) y dos elementos más (LA y RA), adyacentes a cada uno de los subsitios centrales.

2. La formación del complejo nucleoproteico ocurre por la unión ordenada y secuencial de dímeros de proteína a los subsitios del DNA operador. Así, CopG se uniría primero al RSE (sitio por el que CopG presenta la mayor afinidad), después al RA, luego al LSE y finalmente al LA.

3. Las proteínas mutantes $CopG_{G25E}$ y $CopG_{A30E}$ tienen disminuida la afinidad de unión al DNA diana de CopG y afectadas sus propiedades cooperativas. Así, mientras que $CopG_{WT}$ forma cuatro complejos bien definidos cuando se une al operador entero, $CopG_{G25E}$ sólo forma el complejo CI (un dímero unido al RSE) y $CopG_{A30E}$ genera un complejo CII que corresponde a un tetrámero de proteína unido a la mitad derecha del operador.

4. La no-unión de $CopG_{A30E}$ a los subsitios centrales del operador (LSE y RSE) se debe probablemente a la deficiencia de esta proteína mutante para formar interacciones dímero-dímero en el SE y no a una más baja afinidad por el LSE respecto al RA. La repulsión electrostática creada entre las cadenas laterales cargadas negativamente de los residuos del ácido glutámico en posición 30 y los fosfatos del esqueleto del DNA, podría estar involucrada en este hecho.

5. Por NMR se comprobó que tanto la proteína silvestre como las dos mutantes del represor CopG son dímeros en solución, presentan elementos de estructura secundaria idénticos y una configuración tridimensional muy similar.

6. La estabilidad frente a desnaturalización térmica de las tres especies de CopG depende de la concentración, pero es distinta para cada variante proteica. Así pues, la mutante $CopG_{A30E}$ es más estable que $CopG_{WT}$ frente al incremento de temperatura, y ésta última, a su vez, es más resistente a la desnaturalización por calor que la mutante $CopG_{G25E}$.

7. La afinidad de unión de CopG al DNA operador es específica de secuencia. La secuencia de reconocimiento en el sitio primario de unión podría corresponde al motivo de 6-pb 5'-GTGCAC-3', y proponemos la secuencia 5'-(R)TGCA(Y)-3' como consenso para la unión de alta afinidad de la proteína.

8. También proponemos que en la unión de los cuatro dímeros de CopG al operador entero, la proteína se dispone formando dos tetrámeros interrelacionados unidos a una y otra mitad del operador. Esta organización genera dos superficies inter-diméricas distintas en el complejo núcleo-proteico: una, ubicada en la intersección de ambos tetrámeros, conformada por las interacciones establecidas entre los dímeros unidos a los subsitios LSE y RSE (inter-tetramérica), y la segunda, la que se genera entre los dímeros que integran los tetrámeros unidos a cada mitad del operador (intra-tetramérica).

References

REFERENCES

Acebo, P., M. T. Alda, et al. (1996). "Isolation and characterization of pLS 1 plasmid mutants with increased copy numbers." FEMS Microbiology Letters **140**(1): 85-91.

Acebo, P., M. García de Lacoba, et al. (1998). "Structural features of the plasmid pMV158-encoded transcriptional repressor CopG, a protein sharing similarities with both helix-turn-helix and b-sheet DNA binding proteins." Proteins: Structure, Function, and Genetics **32**(2): 248-261.

Aravind, L., V. Anantharaman, et al. (2005). "The many faces of the helix-turn-helix domain: Transcription regulation and beyond." FEMS Microbiology Reviews **29**(2): 231-262.

Augustus, A. M., P. N. Reardon, et al. (2006). "Structural basis for the differential regulation of DNA by the methionine repressor MetJ." Journal of Biological Chemistry **281**(45): 34269-34276.

Backert, S., K. Meissner, et al. (1997). "Unique features of the mitochondrial rolling circle-plasmid mp1 from the higher plant *Chenopodium album* (L.)." Nucleic Acids Research **25**(3): 582-589.

Bachmann, B. J. (1990). "Linkage map of *Escherichia coli* K-12, edition 8." Microbiological Reviews **54**(2): 130-197.

Balganesh, T. S. and S. A. Lacks (1984). "Plasmid vector for cloning in *Streptococcus pneumoniae* and strategies for enrichment for recombinant plasmids." Gene **29**(1-2): 221-230.

Benanti, E. L. and P. T. Chivers (2007). "The N-terminal arm of the *Helicobacter pylori* Ni2+-dependent transcription factor NikR is required for specific DNA binding." Journal of Biological Chemistry **282**(28): 20365-20375.

Benanti, E. L. and P. T. Chivers (2010). "*Geobacter uraniireducens* NikR displays a DNA binding mode distinct from other members of the NikR family." Journal of Bacteriology **192**(17): 4327-4336.

Berns, K. I. (1990). "Parvovirus replication." Microbiological Reviews **54**(3): 316-329.

Bimboim, H. C. and J. Doly (1979). "A rapid alkaline extraction procedure for screening recombinant plasmid DNA." Nucl. Acids Res. **7**(6): 1513-1523.

Bloom, S. L. and D. B. Zamble (2004). "Metal-selective DNA-binding response of *Escherichia coli* NikR " Biochemistry **43**(31): 10029-10038.

Bonvin, A. M. J. J., H. Vis, et al. (1994). "Nuclear magnetic resonance solution structure of the Arc repressor using relaxation matrix calculations." Journal of Molecular Biology **236**(1): 328-341.

Bowie, J. U. and R. T. Sauer (1989). "Equilibrium dissociation and unfolding of the Arc repressor dimer." Biochemistry **28**(18): 7139-7143.

Bowie, J. U. and R. T. Sauer (1990). "TraY proteins of F and related episomes are members of the Arc and Mnt repressor family." Journal of Molecular Biology **211**(1): 5-6.

Brantl, S. (1994). "The copR gene product of plasmid p1P501 acts as a transcriptional repressor at the essential repR promoter." Molecular Microbiology **14**(3): 473-483.

Breg, J. N., J. van Opheusden, et al. (1990). "Structure of Arc repressor in solution: evidence for a family of beta-sheet DNA-binding proteins." Nature **346**(6284): 586-589.

Brown, B. M., J. U. Bowie, et al. (1990). "Arc repressor is tetrameric when bound to operator DNA." Biochemistry **29**(51): 11189-11195.

Brown, M. and R. Sauer (1993). "Assembly of the Arc repressor-operator complex: cooperative interactions between DNA-bound dimers." Biochemistry **32**(5): 1354-1363.

Bujard, H., R. Gentz, et al. (1987). A T5 promoter-based transcription-translation system for the analysis of proteins in Vitro and in Vivo. Methods in Enzymology, Academic Press. **Volume 155:** 416-433.

Burdett, V. (1980). "Identification of tetracycline-resistant R-plasmids in *Streptococcus agalactiae* (group B)." Antimicrob. Agents Chemother. **18**(5): 753-760.

Carrington, P. E., P. T. Chivers, et al. (2003). "Nickel coordination is regulated by the DNA-bound state of NikR." Nat Struct Mol Biol **10**(2): 126-130.

Costa, M., M. Solà, et al. (2001). "Plasmid transcriptional repressor CopG oligomerises to render helical superstructures unbound and in complexes with oligonucleotides." Journal of Molecular Biology **310**(2): 403-417.

Chattoraj, D. K. (2000). "Control of plasmid DNA replication by iterons: no longer paradoxical." Molecular Microbiology **37**(3): 467-476.

Chivers, P. T. and R. T. Sauer (1999). "NikR is a ribbon-helix-helix DNA-binding protein." Protein Science **8**(11): 2494-2500.

Chivers, P. T. and R. T. Sauer (2000). "Regulation of high affinity nickel uptake in bacteria." Journal of Biological Chemistry **275**(26): 19735-19741.

Chivers, P. T. and R. T. Sauer (2002). "NikR repressor: High-affinity nickel binding to the C-terminal domain regulates binding to operator DNA." Chemistry & biology **9**(10): 1141-1148.

Davidson, B. E. and I. S. Girons (1989). "The *Escherichia coli* regulatory protein MetJ binds to a tandemly repeated 8-bp palindrome." Molecular Microbiology **3**(11): 1639-1648.

de la Hoz, A. B., F. Pratto, et al. (2004). "Recognition of DNA by ω protein from the broad-host range *Streptococcus pyogenes* plasmid pSM19035: analysis of binding to operator DNA with one to four heptad repeats." Nucleic Acids Research **32**(10): 3136-3147.

De Pina, K., V. Desjardin, et al. (1999). "Isolation and characterization of the *nikR* gene encoding a nickel-responsive regulator in *Escherichia coli*." Journal of Bacteriology **181**(2): 670-674.

del Solar, G., P. Acebo, et al. (1995). "Replication control of plasmid pLS1: efficient regulation of plasmid copy number is exerted by the combined action of two plasmid components, CopG and RNA II." Molecular Microbiology **18**(5): 913-924.

del Solar, G., A. de la Campa, et al. (1989). "Purification and characterization of RepA, a protein involved in the copy number control of plasmid pLS1." Nucleic Acids Research **17**(7): 2405-2420.

del Solar, G. and M. Espinosa (1992). "The copy number of plasmid pLS1 is regulated by two *trans*-acting plasmid products: the antisense RNA II and the repressor protein, RepA." Molecular Microbiology **6**(1): 83-94.

del Solar, G. and M. Espinosa (2000). "Plasmid copy number control: an ever-growing story." Molecular Microbiology **37**(3): 492-500.

del Solar, G., R. Giraldo, et al. (1998). "Replication and control of circular bacterial plasmids." Microbiol. Mol. Biol. Rev. **62**(2): 434-464.

del Solar, G., A. M. Hernández-Arriaga, et al. (2002). "A genetically economical family of plasmid-encoded transcriptional repressors involved in control of plasmid copy number." J. Bacteriol. **184**(18): 4943-4951.

del Solar, G., M. Moscoso, et al. (1993). "Rolling circle-replicating plasmids from Gram-positive and Gram-negative bacteria: a wall falls." Molecular Microbiology **8**(5): 789-796.

del Solar, G., J. Pérez-Martín, et al. (1990). "Plasmid pLS1-encoded RepA protein regulates transcription from repAB promoter by binding to a DNA sequence containing a 13-base pair symmetric element." J. Biol. Chem. **265**(21): 12569-12575.

del Solar, G. H., A. Puyet, et al. (1987). "Initiation signals for the conversion of single stranded to double stranded DNA forms in the streptococcal plasmid pLS1." Nucleic Acids Research **15**(14): 5561-5580.

Delany, I., R. Leva, et al. (2005). "*In vitro* analysis of protein-operator interactions of the NikR and Fur metal-responsive regulators of coregulated genes in *Helicobacter pylori*." Journal of Bacteriology **187**(22): 7703-7715.

Dixon, W., J. Hayes, et al. (1991). Hydroxyl radical footprinting. Methods in enzymology. **208:** 380-413.

Dosanjh, N. S., A. L. West, et al. (2009). "*Helicobacter pylori* NikR's Interaction with DNA: A Two-Tiered Mode of Recognition." Biochemistry **48**(3): 527-536.

Dostál, L., F. Pratto, et al. (2007). "Binding of regulatory protein omega from *Streptococcus pyogenes* plasmid pSM19035 to direct and inverted 7-base pair repeats of operator DNA." Journal of Raman Spectroscopy **38**(2): 166-175.

Dower, W. J., J. F. Miller, et al. (1988). "High efficiency transformation of *E.coli* by high voltage electroporation." Nucl. Acids Res. **16**(13): 6127-6145.

Erauso, G., S. Marsin, et al. (1996). "Sequence of plasmid pGT5 from the archaeon *Pyrococcus abyss*i: evidence for rolling-circle replication in a hyperthermophile." Journal of Bacteriology **178**(11): 3232-7.

Espinosa, M., S. Cohen, et al. (2005). Plasmid replication and copy number control. The Horizontal Gene Pool: Bacterial Plasmids and Gene Spread. Thomas, CM. Amsterdam, The Netherlands, Harwood Academic Publishers**:** 1-50.

Estrem, S. T., W. Ross, et al. (1999). "Bacterial promoter architecture: subsite structure of UP elements and interactions with the carboxy-terminal domain of the RNA polymerase α subunit." Genes & Development **13**(16): 2134-2147.

Farías, M. E. and M. Espinosa (2000). "Conjugal transfer of plasmid pMV158: uncoupling of the pMV158 origin of transfer from the mobilization gene *mobM*, and modulation of pMV158 transfer in *Escherichia coli* mediated by IncP plasmids." Microbiology **146**(9): 2259-2265.

Fried, M. and D. M. Crothers (1981). "Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis." Nucleic Acids Research **9**(23): 6505-6525.

Galas, D. J. and A. Schmitz (1978). "DNAase footprinting a simple method for the detection of protein-DNA binding specificity." Nucleic Acids Research **5**(9): 3157-3170.

Garvie, C. W. and S. E. V. Phillips (2000). "Direct and indirect readout in mutant Met repressor operator complexes." Structure (London, England : 1993) **8**(9): 905-914.

Gielow, A., L. Diederich, et al. (1991). "Characterization of a phage-plasmid hybrid (phasyl) with two independent origins of replication isolated from *Escherichia coli*." Journal of Bacteriology **173**(1): 73-79.

Golovanov, A. P., D. Barillà, et al. (2003). "ParG, a protein required for active partition of bacterial plasmids, has a dimeric ribbon–helix–helix structure." Molecular Microbiology **50**(4): 1141-1153.

Gomis-Rüth, F. X., M. Solá, et al. (1998a). "The structure of plasmid-encoded transcriptional repressor CopG unliganded and bound to its operator." The EMBO Journal **17**(24): 7404-7415.

Gomis-Rüth, F. X., M. Solà, et al. (1998b). "Overexpression, purification, crystallization and preliminary X-ray diffraction analysis of the pMV158-encoded plasmid transcriptional repressor protein CopG." FEBS Letters **425**(1): 161-165.

Gruss, A. and S. D. Ehrlich (1989). "The family of highly interrelated single-stranded deoxyribonucleic acid plasmids." Microbiological Reviews **53**(2): 231-241.

Hausinger, R. P. (1987). "Nickel utilization by microorganisms." Microbiological Reviews **51**(1): 22-42.

He, Y.-y., P. G. Stockley, et al. (1996). "*In vitro* evolution of the DNA binding sites of *Escherichia coli* methionine repressor, MetJ." Journal of Molecular Biology **255**(1): 55-66.

Hernández-Arriaga, A. M., T. S. Rubio-Lepe, et al. (2009). "Repressor CopG prevents access of RNA polymerase to promoter and actively dissociates open complexes." Nucl. Acids Res. **37**(14): 4799-4811.

Highlander, S. K. and R. P. Novick (1987). "Plasmid repopulation kinatics in *Staphylococcus aureus*." Plasmid **17**: 210-221.

Hol, W., P. van Dujinen, et al. (1978). "The alpha helix dipole and the properties of proteins." Nature **273**: 443-446.

Horinouchi, S. and B. Weisblum (1982). "Nucleotide sequence and functional map of pC194, a plasmid that specifies inducible chloramphenicol resistance." Journal of Bacteriology **150**(2): 815-825.

Hyre, D. E. and L. D. Spicer (1995). "Thermodynamic evaluation of binding interactions in the methionine repressor system of *Escherichia coli* using isothermal titration calorimetry." Biochemistry **34**(10): 3212-3221.

Khan, S. A. (2005). "Plasmid rolling-circle replication: highlights of two decades of research." Plasmid **53**(2): 126-136.

Khan, S. A., G. K. Adler, et al. (1982). "Functional origin of replication of pT181 plasmid DNA is contained within a 168-base-pair segment." Proceedings of the National Academy of Sciences **79**(15): 4580-4584.

Koepsel, R. R., R. W. Murray, et al. (1985). "The replication initiator protein of plasmid pT181 has sequence-specific endonuclease and topoisomerase-like activities." Proceedings of the National Academy of Sciences **82**(20): 6845-6849.

Lacks, S. (1966). "Integration efficiency and genetic recombination in penumococcal transformation." Genetics **53**(1): 207-235.

Lacks, S. and B. Greenberg (1977). "Complementary specificity of restriction endonucleases of *Diplococcus pneumoniae* with respect to DNA methylation." Journal of Molecular Biology **114**(1): 153-168.

Lacks, S. A., P. López, et al. (1986). "Identification and analysis of genes for tetracycline resistance and replication functions in the broad-host-range plasmid pLS1." Journal of Molecular Biology **192**(4): 753-765.

López-Aguilar, C., J. A. Ruiz-Masó, et al. (2013). "Translation initiation of the replication initiator repB gene of promiscuous plasmid pMV158 is led by an extended non-SD sequence." Plasmid **70**(1): 69-77.

Lorenzo-Díaz, F. and M. Espinosa (2009). "Large-scale filter mating assay for intra- and inter-specific conjugal transfer of the promiscuous plasmid pMV158 in Gram-positive bacteria." Plasmid **61**(1): 65-70.

Lu, J., A. den Dulk-Ras, et al. (2009). "*Agrobacterium tumefaciens* VirC2 enhances T-DNA transfer and virulence through its C-terminal ribbon-helix-helix DNA-binding fold." Proceedings of the National Academy of Sciences **106**(24): 9643-9648.

Lum, P. L., M. E. Rodgers, et al. (2002). "TraY DNA Recognition of Its Two F Factor Binding Sites." Journal of Molecular Biology **321**(4): 563-578.

Lum, P. L. and J. F. Schildbach (1999). "Specific DNA recognition by F factor TraY involves b-sheet residues." Journal of Biological Chemistry **274**(28): 19644-19648.

Luria, S. E. and J. W. Burrous (1957). "Hybridization between *Escherichia coli* and *Shigella*." J. Bacteriol. **74**(4): 461-476.

Luscombe, N. M., R. A. Laskowski, et al. (2001). "Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level." Nucleic Acids Research **29**(13): 2860-2874.

Man, T.-K. and G. D. Stormo (2001). "Non-independence of Mnt repressor-operator interaction determined by a new quantitative multiple fluorescence relative affinity (QuMFRA) assay." Nucleic Acids Research **29**(12): 2471-2478.

Mao, C., N. G. Carlson, et al. (1994). "Cooperative DNA-protein Interactions: Effects of changing the spacing between adjacent binding sites." Journal of Molecular Biology **235**(2): 532-544.

Marincs, F., I. W. Manfield, et al. (2006). "Transcript analysis reveals an extended regulon and the importance of protein-protein co-operativity for the *Escherichia coli* methionine repressor." Biochem J **396**(2): 227-234.

Maroney, M. J. (1999). "Structure/function relationships in nickel metallobiochemistry." Current Opinion in Chemical Biology **3**(2): 188-199.

Matthews, B. W., D. H. Ohlendorf, et al. (1982). "Structure of the DNA-binding region of lac repressor inferred from its homology with cro repressor." Proceedings of the National Academy of Sciences **79**(5): 1428-1432.

Maxam, A. and W. Gilbert (1980). "Sequencing end-labeled DNA with base-specific chemical cleavages." Methods Enzymol **65**(1): 499-560.

Minton, A. P. (1994). Conservation of signal:a new algorithm for the elimination of the reference concentration as an independently variable para meter in the analysis of sedimentation equilibrium. Birkhauser, Boston, MA.

Mobley, H. L., M. D. Island, et al. (1995). "Molecular biology of microbial ureases." Microbiological Reviews **59**(3): 451-80.

Mulrooney, S. B. and R. P. Hausinger (2003). "Nickel uptake and utilization by microorganisms." FEMS Microbiology Reviews **27**(2-3): 239-261.

Navarro, C., L.-F. Wu, et al. (1993). "The nik operon of *Escherichia coli* encodes a periplasmic binding-protein-dependent transport system for nickel." Molecular Microbiology **9**(6): 1181-1191.

Nieto, C. and M. Espinosa (2003). "Construction of the mobilizable plasmid pMV158GFP, a derivative of pMV158 that carries the gene encoding the green fluorescent protein." Plasmid **49**(3): 281-285.

Nordström, K. and E. G. Wagner (1994). "Kinetic aspect of control of plasmid replication by antisense RNA." Trends in biochemical sciences **19**(7): 294-300.

Novick, R. P. (1989). "Staphylococcal plasmids and their replication." Annual Review of Microbiology **43**(1): 537-563.

Old, I. G., S. E. V. Phillips, et al. (1991). "Regulation of methionine biosynthesis in the enterobacteriaceae." Progress in Biophysics and Molecular Biology **56**(3): 145-185.

Old, I. G., I. Saint-Girons, et al. (1993). "Physical mapping of the scattered methionine genes on the *Escherichia coli* chromosome." Journal of Bacteriology **175**(11): 3689-3691.

Overgaard, M., J. Borch, et al. (2009). "RelB and RelE of *Escherichia coli* form a tight complex that represses transcription via the Ribbon-Helix-Helix motif in RelB." Journal of Molecular Biology **394**(2): 183-196.

Pabo, C. O. and R. T. Sauer (1984). "Protein-DNA recognition." Annual Review of Biochemistry **53**(1): 293-321.

Parsons, I. D., B. Persson, et al. (1995). "Probing the molecular mechanism of action of co-repressor in the *E.coli* methionine repressor-operator complex using surface plasmon resonance (SPR)." Nucleic Acids Research **23**(2): 211-216.

Pérez-Martín, J., G. H. del Solar, et al. (1989). "Induced bending of plasmid pLS1 DNA by the plasmid-encoded protein RepA." J. Biol. Chem. **264**(35): 21334-21339.

Philo, J. S. (1997). "An improved function for fitting sedimentation velocity data for low-molecular-weight solutes." Biophysical Journal **72**(1): 435-444.

Phillips, C. M., P. S. Nerenberg, et al. (2009). "Physical basis of metal-binding specificity in *Escherichia coli* NikR." Journal of the American Chemical Society **131**(29): 10220-10228.

Phillips, C. M., E. R. Schreiter, et al. (2008). "Structural basis of the metal specificity for nickel regulatory protein NikR." Biochemistry **47**(7): 1938-1946.

Phillips, C. M., E. R. Schreiter, et al. (2010). "Structural basis of low-affinity nickel binding to the nickel-responsive transcription factor NikR from *Escherichia coli*." Biochemistry **49**(36): 7830-7838.

Phillips, K. and S. E. V. Phillips (1994). "Electrostatic activation of *Escherichia coli* methionine repressor." Structure **2**(4): 309-316.

Phillips, S. E. V., I. Manfield, et al. (1989). "Cooperative tandem binding of met repressor of *Escherichia coli*." Nature **341**(6244): 711-715.

Priebe, S. D. and S. A. Lacks (1989). "Region of the streptococcal plasmid pMV158 required for conjugative mobilization." J. Bacteriol. **171**(9): 4778-4784.

Pritchard, R. H., P. T. Barth, et al. (1969). "Control of DNA synthesis in bacteria." Symp. Soc. Gen. Microbiol. **19**: 263-297.

Rafferty, J. B., W. S. Somers, et al. (1989). "Three-dimensional crystal structures of *Escherichia coli* Met repressor with and without corepressor." Nature **341**(6244): 705-710.

Ramsey, D. M., P. J. Baynham, et al. (2005). "Binding of *Pseudomonas aeruginosa* AlgZ to sites upstream of the algZ promoter leads to repression of transcription." Journal of Bacteriology **187**(13): 4430-4443.

Rasooly, A. and R. P. Novick (1993). "Replication-specific inactivation of the pT181 plasmid initiator protein." Science **262**(5136): 1048-1050.

Raumann, B. E., M. A. Rould, et al. (1994). "DNA recognition by b-sheets in the Arc represser-operator crystal structure." Nature **367**(6465): 754-757.

Rippe, K. (1997). "Analysis of protein-DNA binding at equilibrium." B.I.F. Futura **12**: 20-26.

Rojo, F. (1999). "Repression of transcription initiation in bacteria." Journal of Bacteriology **181**(10): 2987-2991.

Rojo, F. (2001). "Mechanisms of transcriptional repression." Current Opinion in Microbiology **4**(2): 145-151.

Ross, W., A. Ernst, et al. (2001). "Fine structure of *E. coli* RNA polymerase-promoter interactions: α subunit binding to the UP element minor groove." Genes & Development **15**(5): 491-506.

Rowe, J. L., G. L. Starnes, et al. (2005). "Complex transcriptional control links NikABCDE-dependent nickel transport with hydrogenase expression in *Escherichia coli*." Journal of Bacteriology **187**(18): 6317-6323.

Rozhon, W., M. Khan, et al. (2011). "Identification of *cis*- and *trans*-acting elements in pHW126, a representative of a novel group of rolling circle plasmids." Plasmid **65**(1): 70-76.

Ruiz-Masó, J. A., R. Lurz, et al. (2007). "Interactions between the RepB initiator protein of plasmid pMV158 and two distant DNA regions within the origin of replication." Nucleic Acids Research **35**(4): 1230-1244.

Saint-Girons, I., N. Duchange, et al. (1984). "Structure and autoregulation of the metJ regulatory gene in *Escherichia coli*." Journal of Biological Chemistry **259**(22): 14282-5.

Sambrook, J., E. F. Fritsch, et al. (1989). Molecular Cloning: A laboratory manual, Cold Spring Harbor Laboratory Press N.Y.

Sauer, R. T., W. Krovatin, et al. (1983). "Primary structure of the immI immunity region of bacteriophage P22." Journal of Molecular Biology **168**(4): 699-713.

Schildbach, J. F., A. W. Karzai, et al. (1999). "Origins of DNA-binding specificity: Role of protein contacts with the DNA backbone." Proceedings of the National Academy of Sciences **96**(3): 811-817.

Schreiter, E. R. and C. L. Drennan (2007). "Ribbon-helix-helix transcription factors: variations on a theme." Nature Reviews. Microbiology **5**(9): 710-720.

Schreiter, E. R., M. D. Sintchak, et al. (2003). "Crystal structure of the nickel-responsive transcription factor NikR." Nat Struct Mol Biol **10**(10): 794-799.

Schreiter, E. R., S. C. Wang, et al. (2006). "NikR-operator complex structure and the mechanism of repressor activation by metal ions." Proceedings of the National Academy of Sciences **103**(37): 13676-13681.

Schuck, P. (2000). "Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling." Biophysical Journal **78**(3): 1606-1619.

Senear, D. F. and M. Brenowitz (1991). "Determination of binding constants for cooperative site-specific protein-DNA interactions using the gel mobility-shift assay." Journal of Biological Chemistry **266**(21): 13661-71.

Smith, M., N. Shoemaker, et al. (1980). "Transfer of plasmid by conjugation in *Streptococcus pneumoniae*." Plasmid **3**(1): 70-79.

Smith, T. L. and R. T. Sauer (1995). "P22 Arc Repressor: Role of cooperativity in repression and binding to operators with altered half-site spacing." Journal of Molecular Biology **249**(4): 729-742.

Smith, T. L. and R. T. Sauer (1996). "Role of operator subsites in Arc repression." Journal of Molecular Biology **264**(2): 233-242.

Somers, W. S. and S. E. V. Phillips (1992). "Crystal structure of the met repressor-operator complex at 2.8 Å resolution reveals DNA recognition by B-strands." Nature **359**(6394): 387-393.

Somers, W. S., J. B. Rafferty, et al. (1994). "The Met repressor-operator complex: DNA recognition by β-Strands." Annals of the New York Academy of Sciences **726**(1): 105-117.

Spector, S., R. T. Sauer, et al. (2004). "Computational and experimental probes of symmetry mismatches in the Arc repressor-DNA complex." Journal of Molecular Biology **340**(2): 253-261.

Stassi, D. L., P. Lopez, et al. (1981). "Cloning of chromosomal genes in Streptococcus pneumoniae." Proceedings of the National Academy of Sciences of the United States of America **78**(11): 7028-7032.

Stingl, K. and H. De Reuse (2005). "Staying alive overdosed: How does Helicobacter pylori control urease activity?" International Journal of Medical Microbiology **295**(5): 307-315.

te Riele, H., B. Michel, et al. (1986). "Are single-stranded circles intermediates in plasmid DNA replication?" EMBO J. **5**(3): 631-637.

Tullius, T. D. and B. A. Dombroski (1986). "Hydroxyl radical "footprinting": high-resolution information about DNA-protein contacts and application to lambda repressor and Cro protein." Proceedings of the National Academy of Sciences **83**(15): 5469-5473.

van Vliet, A. H. M., F. D. Ernst, et al. (2004). "NikR-mediated regulation of *Helicobacter pylori* acid adaptation." Trends in Microbiology **12**(11): 489-494.

Vershon, A. K., R. D. Kelley, et al. (1989). "Sequence-specific binding of arc repressor to DNA. Effects of operator mutations and modifications." Journal of Biological Chemistry **264**(6): 3267-3273.

Vershon, A. K., S.-M. Liao, et al. (1987a). "Bacteriophage P22 Mnt repressor: DNA binding and effects on transcription in vitro." Journal of Molecular Biology **195**(2): 311-322.

Vershon, A. K., S.-M. Liao, et al. (1987b). "Interaction of the bacteriophage P22 arc repressor with operator DNA." Journal of Molecular Biology **195**(2): 323-331.

Vershon, A. K., P. Youderian, et al. (1985). "The bacteriophage P22 arc and mnt repressors. Overproduction, purification, and properties." Journal of Biological Chemistry **260**(22): 12124-9.

Villarejo, M. R. and I. Zabin (1974). "β-Galactosidase from termination and deletion mutant strains." Journal of Bacteriology **120**(1): 466-474.

Wang, S. C., A. V. Dias, et al. (2004). "Selectivity of metal binding and metal-induced stability of *Escherichia coli* NikR " Biochemistry **43**(31): 10018-10028.

Wang, S. C., Y. Li, et al. (2010). "Potassium is critical for the Ni(II)-responsive DNA-binding activity of *Escherichia coli* NikR." Journal of the American Chemical Society **132**(5): 1506-1507.

Wei, Y. and E. B. Newman (2002). "Studies on the role of the metK gene product of *Escherichia coli* K-12." Molecular Microbiology **43**(6): 1651-1656.

Weihofen, W. A., A. Cicek, et al. (2006). "Structures of ω repressors bound to direct and inverted DNA repeats explain modulation of transcription." Nucleic Acids Research **34**(5): 1450-1458.

Williams, S. L. and J. F. Schildbach (2007). "TraY and Integration Host Factor *oriT* binding sites and F conjugal transfer: sequence variations, but not altered spacing, are tolerated." Journal of Bacteriology **189**(10): 3813-3823.

Wishart, D. S. and B. D. Sykes (1994). "The ^{13}C chemical-shift index: A simple method for the identification of protein secondary structure using ^{13}C chemical-shift data." Journal of Biomolecular NMR **4**(2): 171-180.

Wu, L.-F., C. Navarro, et al. (1991). "The hydC region contains a multi-cistronic operon (nik) involved in nickel transport in *Escherichia coli*." Gene **107**(1): 37-42.

Wu, L. F., M. A. Mandrand-Berthelot, et al. (1989). "Nickel deficiency gives rise to the defective hydrogenase phenotype of hydc and fnr mutants in *Escherichia coli*." Molecular Microbiology **3**(12): 1709-1718.

Wyman, J. (1964). "Linked functions and reciprocal effects in hemoglobin: a second look." Adv Protein Chem **19**: 223-286.

Yasukawa, H., T. Hase, et al. (1991). "Rolling-circle replication of the plasmid pKYM isolated from a gram-negative bacterium." Proceedings of the National Academy of Sciences **88**(22): 10282-10286.

Related publications

Published online 11 June 2009

Nucleic Acids Research, 2009, Vol. 37, No. 14 ***4799–4811***
doi:10.1093/nar/gkp503

Repressor CopG prevents access of RNA polymerase to promoter and actively dissociates open complexes

Ana M. Hernández-Arriaga, Tania S. Rubio-Lepe, Manuel Espinosa and Gloria del Solar*

Centro de Investigaciones Biológicas, Consejo Superior de Investigaciones Científicas, Madrid, Spain

Received February 15, 2009; Revised May 12, 2009; Accepted May 23, 2009

ABSTRACT

Replication of the promiscuous plasmid pMV158 requires expression of the initiator *repB* gene, which is controlled by the repressor CopG. Genes *repB* and *copG* are co-transcribed from promoter P_{cr}. We have studied the interactions between RNA polymerase, CopG and the promoter to elucidate the mechanism of repression by CopG. Complexes formed at 0°C and at 37°C between RNA polymerase and P_{cr} differed from each other in stability and in the extent of the DNA contacted. The 37°C complex was very stable (half-life of about 3 h), and shared features with typical open complexes generated at a variety of promoters. CopG protein repressed transcription from P_{cr} at two different stages in the process leading to the initiation complex. First, CopG hindered binding of RNA polymerase to the promoter. Second, CopG was able to displace RNA polymerase once the enzyme has formed a stable complex with P_{cr}. A model for the CopG-mediated disassembly of the stable RNA polymerase–P_{cr} promoter complex is presented.

INTRODUCTION

Inhibition of transcription initiation in bacteria can take place at any of the steps in the process leading to the formation of the elongation complex (1,2). Although some of the best-characterized transcriptional repressors seem to act by hindering the access of the RNA polymerase (RNAP) to the promoter, an increasing number of regulatory proteins have been reported that inhibit transcription initiation at a step subsequent to RNAP binding. Thus, the repressors λ*c*I (3), LexA (4), LacI (5,6) or MalR (7) occlude binding of RNAP to their respective promoters (P_R, P_{uvrA}, P_{lac}, or P_M), while GalR bound to operator O_E inhibits isomerization from the ternary closed complex to the open complex at promoter *galP*1 (8,9), and phage ϕ29 protein p4 prevents RNAP from leaving the A2c promoter (10). Inhibition of transcription at a step subsequent to promoter melting has also been shown for some of the reported regulatory activities of the architectural proteins H-NS and FIS. Thus, binding of FIS to two sites centred at positions −62 and −109 with respect to the transcription start point of the *gyrB* gene does not interfere with generation of an open complex at the *gyrB* promoter, but seems to prevent the NTP-driven isomerization of open complexes to initiation complexes (11). The presence of H-NS, while stimulating open complex formation at the rRNA *rrnB* P1 promoter of *Escherichia coli*, dramatically reduces the generation of transcription products longer than 3 nt from this promoter (12). An interesting example of growth phase-dependent regulation of gene expression by H-NS is the selective repression of the transcription initiated by the RNAP σ^{70} holoenzyme at promoter *hdeAB* (13). RNAP holoenzymes carrying either the house-keeping σ subunit (σ^{70}) or the stationary phase σ (σ^{38}) can form an open complex at the *hdeAB* promoter in the presence of H-NS, although the RNAP σ^{70} holoenzyme is unable to initiate transcription. This selective repression is attributed to differences in the degree of DNA wrapping around the two holoenzymes: the tighter wrapping around RNAP σ^{70} holoenzyme allows extension of H-NS by cooperative binding from its AT-tract nucleation site centred at position −118 to promoter-downstream sequences, which results in effective sealing of the DNA loop and trapping of RNAP (13).

The mechanism by which a given repressor acts can condition the relative location of its DNA target with respect to the promoter, and vice versa. Hence, the step at which transcription initiation is inhibited differs between the closely-related repressors of coliphages λ and 434 (14). These two proteins share both a common structure and the ability to bind specifically to a set of three operator sites (O_R1, O_R2 and O_R3) located between two divergently oriented promoters, P_R and P_{RM}. In both phages, the region of the DNA containing O_R1 and O_R2 (to which the repressors bind with the highest affinity) overlaps the region containing the –35 and –10 boxes

*To whom correspondence should be addressed. Tel: +34 91 837 3112; Fax: +34 91 536 0432; Email: gdelsolar@cib.csic.es

of P_R, although the relative arrangement of the operator sites with respect to the promoter elements differs in λ and 434 (14). This distinct architecture has been invoked to argue the different mechanism of repression of the P_R activity found in each of these phages. In λ, where O_R1 and O_R2 partially overlap the –10 and –35 elements of P_R respectively, the binding of the repressor to either operator site blocks the access of RNAP to the promoter. In 434, where O_R1 and O_R2 are immediately downstream from the –10 element and the –35 element of P_R respectively, binding of the repressor to O_R2 is necessary and sufficient to inhibit transcription from P_R through a mechanism that does not exclude the subsequent binding of RNAP. In this case, a stable (heparin-resistant) ternary complex is generated, which cannot progress to an open complex (14). From these latter examples, it becomes clear that prediction of the inhibitory mechanism of transcription initiation should be based on the relative arrangement and topology of RNAP and repressor on the DNA double helix, rather than in sequence or structural similarities to a reference protein acting at a known step of the initiation process.

Transcriptional repressor CopG, encoded by the promiscuous plasmid pMV158, is a small protein (45 amino-acid polypeptide chains) which, by inhibiting synthesis of the *cop-rep* mRNA, regulates both the expression of its own gene and that of the initiator of replication (*repB*) gene. Contacts of CopG to its target DNA span about 50 bp, through a region that includes the –35 and –10 boxes of the regulated P_{cr} promoter (15) (Figure 1A). In the centre of the contacted DNA, and overlapping the –35 box, there is a 13-bp pseudo-symmetric element (SE; Figure 1A) which constitutes the primary target of CopG. Although the structure solved from the co-crystals of CopG bound to either a 19- or a 22-bp double-stranded DNA shows two protein dimers, each interacting with a half of the SE, the working model for the whole DNA specifically contacted by CopG assumes the cooperative binding of four dimers of the protein (16,17). CopG belongs to the ribbon-helix-helix (RHH) class of DNA-binding proteins, which interact specifically with the bases of the DNA through residues located in the two-stranded antiparallel β-sheet (18,19). Members of the RHH class exhibit cooperativity based on protein–protein interactions, generating at least a dimer of protein dimers upon binding to their target DNA (20–22). CopG, which is no more than the RHH motif, is the smallest among these proteins and represents the minimal DNA-binding structure within this class of proteins (17). The peptide-backbone structure of the CopG dimer is almost identical to that of the RHH motif of Arc repressor from *Salmonella* bacteriophage P22, although this latter protein is slightly bigger (53 residues polypeptide chains) and contains an additional N-terminal region which also interacts with the target DNA (17,19). The similarity between the Arc- and CopG-mediated transcriptional regulatory systems also extends to the DNA moiety. In both systems, the operator to which the repressor binds overlaps the promoter region and contains two inversely-repeated copies of a 4-bp sequence (5′-TAGA-3′ for Arc operator, and the self-palindrome 5′-TGCA-3′ for CopG operator) which are 5 bp apart. This arrangement results in 13-bp SEs, whose dyad axes pass in both operators through a central G:C base pair. A dimer of either Arc or CopG interacts with each half of the respective SE establishing direct contacts with at least one base of each of the base pairs constituting the 4-bp inversely-repeated boxes, and with two further bases outside the 13-bp palindromic element (one at each side of the SE) (16,19). With respect to the mechanism of transcription inhibition, some detailed information has been published for the Arc-regulated system. Binding of Arc to its operator represses transcription from two divergent and overlapping promoters, P_{ant} and P_{mnt} (23). It has been shown that binding of Arc does not preclude RNAP binding to P_{ant}, but prevents the polymerase from forming a heparin-resistant stable complex at this promoter, so that the rate of generation of a transcriptionally-competent open complex is reduced (23). Although CopG-mediated repression of transcription from P_{cr} has been reported both *in vivo* and *in vitro* (15), the mechanisms of this repression have not been investigated previously. Here we characterize the complexes formed by binding of RNAP to the P_{cr} promoter at 0°C and 37°C. In addition, we present results showing that CopG does not only prevent binding of RNAP to the promoter, but is also able to dislodge the polymerase once it has formed a stable heparin-resistant complex.

MATERIALS AND METHODS

DNA, proteins and chemicals

A 239-bp DNA fragment containing P_{cr} and the overlapping CopG target was generated by PCR using Pfu DNA polymerase (Stratagene), a pair of synthetic primers (5′-CGCCTTTAGCCTTAGAG-3′ and 5′-CCATCTCTC TTGCCAT-3′), and pMV158 DNA as template. To label the DNA fragment at only one 5′-end, either of these primers was treated with T4 polynucleotide kinase (New England Biolabs) and [γ-^{32}P]ATP (Amersham) before performing the amplification reaction. The amplified fragment spanned from −144 to +95 relative to the P_{cr} transcription start site (15). This fragment was used in all the experiments described below. DNase I and *E. coli* RNAP holoenzyme were from Boehringer Mannheim. When indicated, RNAP holoenzyme from USB was used. Promoter binding activity of the commercial preparations of RNAP was determined by EMSA, using DNA concentrations 10, 20 or 40 times the total concentration of RNAP (4 nM). RNAP holoenzymes from Boehringer and USB were ∼15–20% and ∼65% active, respectively; concentrations refer to active holoenzyme. CopG was purified as described (24), and dialyzed against binding buffer (25 mM Tris–HCl pH 8.0, 1.5 mM EDTA, 6 mM DTT, 170 mM KCl, 5 mM $MgCl_2$, 2.5 mM $CaCl_2$, 7.5% glycerol and 5% ethylene glycol) to avoid changing the salt concentration when the protein is added to a binding mixture. Concentrations of CopG given throughout this article refer to total protomers. All other reagents were of the highest commercially available grade.

Formation and detection of protein–DNA complexes

Reactions were performed in 5 µl of the BB buffer (binding buffer plus 500 µg/ml bovine serum albumin). RNAP–DNA or CopG–DNA complexes were formed by mixing 2 nM (labelled) or 10 nM (unlabelled) DNA fragment with varying concentrations of either protein. RNAP–DNA-binding reactions were incubated either at 37°C for 30 min or at 0°C for 60 min. When indicated, RNAP–DNA complexes were treated with heparin (150 µg/ml final concentration, unless stated otherwise) for 5 min at the same temperature, to eliminate short-lived complexes. Treatment with the polyanion was omitted when binding was performed at 0°C, to avoid removal of the unstable complexes that usually result at this temperature. CopG–DNA mixtures were incubated at 37°C or 0°C for 10 min. Under the above conditions, binding mixtures reached equilibrium. Samples were loaded on running 5% PAA gels. The same temperature was maintained throughout the whole EMSA. Labelled DNA bands were detected by autoradiography and quantified with the storage phosphor technology, with the aid of a FLA-3000 (FUJIFILM) imaging system and the Quantity One software (Bio-Rad).

DNase I footprinting

Binding reactions (50 µl) contained 2 nM labelled DNA in BB buffer. Proteins CopG and/or RNAP were added at the concentrations indicated. Binding of RNAP was carried out at 37°C for 30 min, or at 0°C for 60 min. CopG binding was performed at 0°C or 37°C, for 10 min. Probing with DNase I was carried out at the same temperature as the binding of CopG and/or RNAP. Digestion (3 min) of the DNA was started by adding 2×10^{-5} (37°C) or 7×10^{-3} (0°C) units of DNase I. Samples were processed as described (15).

HO• footprinting

RNAP–DNA or CopG–DNA complexes were generated at 37°C in 100 µl of BB buffer. Samples were dialyzed against 8 mM Tris–HCl pH 8.0, for 60 min at 37°C to remove the glycerol, and then treated with heparin for 5 min at the same temperature. Although dialysis against low-salt buffer has been reported to stabilize the RNAP–promoter complexes (25), no differences in complex stability or in the footprinting pattern were observed when samples were dialyzed against 25 mM Tris–HCl pH 8.0 and 170 mM KCl. To cleave the DNA backbone, reagents for HO• production [$Fe(EDTA)^{2-}$, H_2O_2 and sodium ascorbate] were added in two alternative ways: method 1, a freshly prepared 6×-concentrated mixture that gives final concentrations of 0.1 mM Fe (II), 0.2 mM Na_2 EDTA, 0.3% H_2O_2 and 1 mM sodium ascorbate; or method 2, freshly made solutions which were added separately in the following order: sodium ascorbate to 1 mM, H_2O_2 to 0.03% and Fe $(EDTA)^{2-}$ [final concentrations of 9 mM for Fe (II) and 18 mM for Na_2 EDTA]. Cleavage with HO• was allowed for 2 min at 37°C and stopped by addition of a thiourea/EDTA solution (final concentrations 9.5 mM thiourea and 1.7 mM EDTA). The treatment described as method 1, while not destabilizing the CopG–DNA complexes, disrupts the RNAP bound to the DNA so that only ~30% of the initial fraction of complexes was observed after the 2 min HO• treatment. To obtain a clear RNAP-mediated footprinting pattern, RNAP–DNA complexes were next separated from free DNA by electrophoresis on native 5% PAA gels. The DNA of the complex was eluted and precipitated before loading onto the sequencing gel. When the RNAP binding mixtures were probed with HO• under the conditions described as method 2, more than 90% of the total DNA remained complexed to RNAP after the 2 min treatment, making isolation of RNAP–DNA complexes unnecessary, as was the case with the CopG binding mixtures treated as in method 1. DNA was analyzed on 8% PAA sequencing gels. G + A sequencing reactions were performed as described by (26).

Stability of the RNAP–DNA and CopG–DNA complexes

Equilibrium mixtures were prepared by incubating 2 nM labelled DNA fragment and varying concentrations of protein (RNAP or CopG) in BB buffer, at either 37°C or 0°C. Dissociation of complexes between labelled DNA and either protein was made irreversible by addition of the indicated molar excess of competing unlabelled promoter fragment. Samples were taken at intervals and applied directly to running gels. Fractions of complexed- and free-labelled DNA were directly quantified as above. Data were analyzed according to the equation for a first-order kinetic process:

$$\mathrm{Ln}\left(\frac{[\mathrm{Protein}-\mathrm{DNA}]}{[\mathrm{Protein}-\mathrm{DNA}]_0}\right) = -k_d t, \quad \mathbf{1}$$

where [Protein-DNA] represents the concentration of the complex between either RNAP or CopG and the labelled DNA at time t, and $[\text{Protein-DNA}]_0$ is its value at $t = 0$. Dissociation rate constants of the complexes ($k_d \pm 2$ SE) were estimated from least-squares analysis of the data. The half-life ($t_{1/2}$) of the complexes was calculated according to the following equation:

$$t_{1/2} = \frac{\mathrm{Ln}2}{k_d}. \quad \mathbf{2}$$

RESULTS

Analysis of the DNA backbone regions of promoter P_{cr} contacted by RNAP at 0°C and 37°C

Isomerization of a short-lived closed complex to a stable open complex is usually accompanied by extension of the enzyme contacts to regions downstream of the transcription start site (position +1) (25,27). In addition, kinetic studies have evidenced that unstable closed complexes accumulate at binding temperatures below 10°C, whereas the open complex predominates above 21°C (25,28,29). However, the characteristic DNA-protection pattern of the different complexes that can be 'trapped' and analyzed has to be investigated for each given promoter, as several examples differing from the above general scheme have

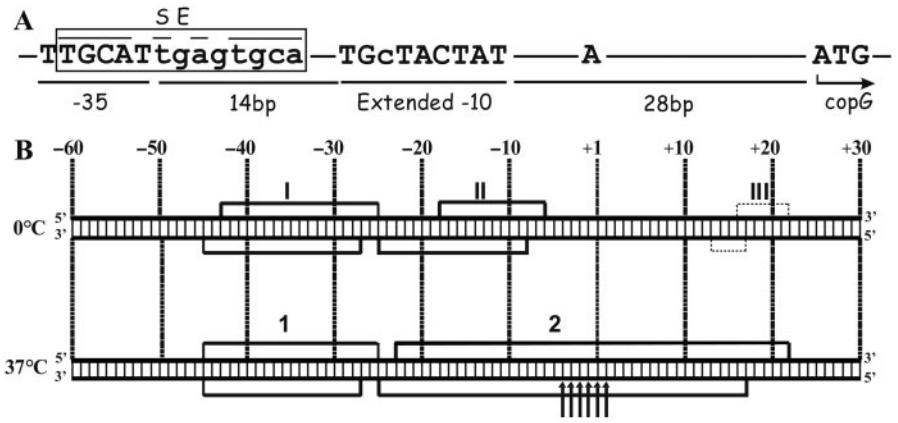


Figure 1. The P_{cr} promoter of pMV158. (**A**) Conserved elements of P_{cr}. The *copG* start codon, the +1 site, and the −35 and extended −10 elements are in capital letters; distances between them are shown. The SE of the CopG operator is boxed. (**B**) Footprinting pattern of the 0°C and 37°C RNAP–P_{cr} complexes. The scheme summarizes the results shown in Supplementary Figures S1 and S2. Promoter positions are numbered relative to the transcription start site. DNase I protections are denoted by numbered brackets. Thin dotted-line brackets indicate weak protections. Enhancements to HO• attack are shown by arrows.

been reported. Thus, a short-lived closed complex that displays RNAP contacts spanning ~30 bases downstream of the +1 site is generated at 0°C between σ^{70} RNAP and the λP_R promoter (30). Also, the downstream boundary of the DNase I footprint of the complex formed at 4°C between σ^{70} RNAP and the phage λ *prmup-1* Δ265 promoter is around position +10 (31).

The contact pattern on both DNA strands of the binary complexes generated upon binding of RNAP to the pMV158 P_{cr} promoter at 0°C and at 37°C were analyzed by chemical (hydroxyl radical, HO•) and enzymatic (DNase I) probing of the DNA backbone reactivity. At 0°C, DNase I footprinting revealed strong RNAP-mediated protections extending from –43 to –25 and from –18 to –6 on the non-template strand, and, on the template strand, from –45 to –27 and from –25 to –8. In addition to these main footprints, a weak protection against DNase I cleavage was observed downstream of the transcription start site (from +16 to +22 on the non-template strand, and from +13 to +17 on the template strand; Figure 1B and Supplementary Figure S1). The different degree of protection observed between the upstream and downstream footprints suggests that the binding equilibrium at 0°C is heterogeneously populated. In the main complex, the overall protected area would only span upstream of the start site (between positions –45 and –6), yielding a footprint pattern similar to those observed for low-temperature complexes at a variety of promoters (typically from positions –55 to –5). The weakly protected area could arise from a small fraction of complexes in which the RNAP contacts extend further downstream, reaching around +20. The presence of a mixed population of RNAP–promoter complexes at equilibrium at a temperature as low as 0°C seems to be very uncommon, and has only been reported for the *E. coli* σ^{32} RNAP–*groE* promoter binary complex (32). At 37°C, the same average degree of protection was observed upstream and downstream of the +1 site, with strong footprints spanning from –45 to –25 and from –23 to +22 on the non-template strand, and from –45 to –27 and from –25 to +17 on the template strand (Figure 1B and Supplementary Figure S1). Thus, at 37°C the overall protected area extended further downstream by about 25 nt relative to that in the predominant complex generated at 0°C.

Additional information on the conformation of the RNAP–P_{cr} binary complex generated at 37°C was achieved by HO• probing of the reactivity of the DNA backbone of the template strand around the +1 site. High reactivity of the sequence between positions –7 and +2 of only the template strand has been observed in the *E. coli* RNAP–T7 A1 promoter complex generated at 37°C, whereas the same region is protected in the 20°C complex and, to a lesser extent, in the complex resulting at 30°C (25). This strand-specific, temperature-dependent sensitivity of the +1 region towards HO• has been proposed to characterize the open complexes by distinguishing them from intermediate closed complexes that otherwise have the same upstream and downstream boundaries of the overall protected region (25). After binding of RNAP to P_{cr} at 37°C and selection of the stable binary complexes by a brief treatment with heparin, the DNA was cleaved with HO•. Hypersensitive sites, surrounded by protected regions, were clearly visualized between positions –4 and +2 of the template strand (Figure 1B and Supplementary Figure S2), suggesting that at least a significant fraction of the complexes generated at 37°C are in an open conformation.

Stability of the binary RNAP–P_{cr} complexes generated at 0°C and 37°C

We analyzed the kinetics of dissociation of the short and extended RNAP–P_{cr} complexes generated at 0°C and 37°C, respectively. Equilibrium mixtures containing 2 nM-labelled promoter fragment and 150 nM RNAP were prepared at each temperature (under these conditions, essentially all P_{cr} promoters are bound by the polymerase). Dissociation was initiated by addition of a

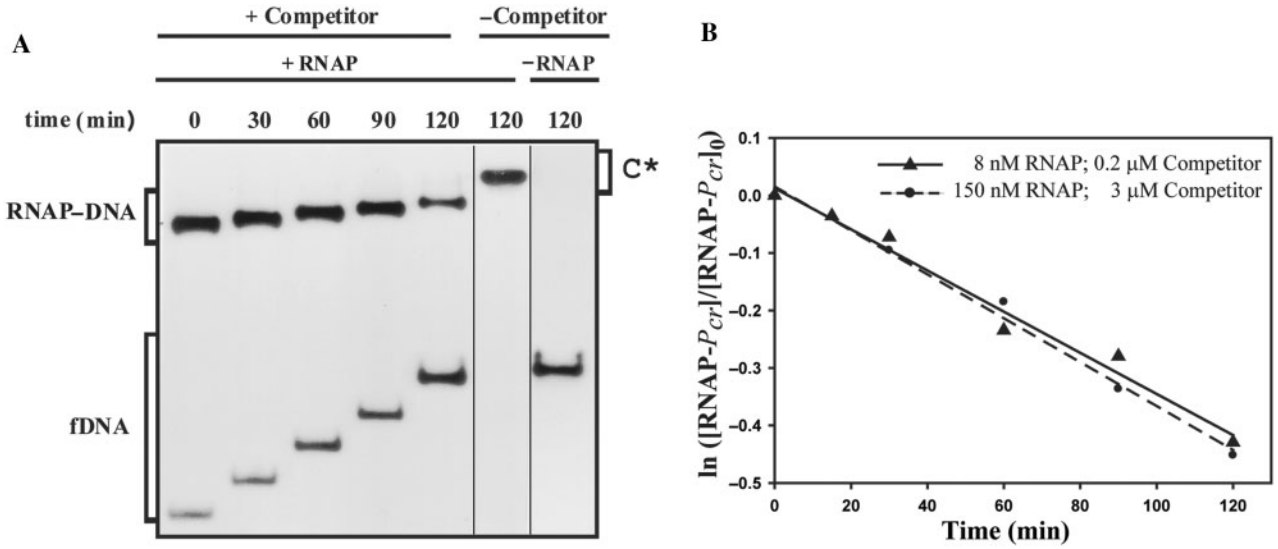


Figure 2. Kinetic of dissociation of RNAP from P_{cr} at 37°C. (**A**) EMSA analysis of the stability of the RNAP–P_{cr} complexes. Equilibrium mixtures contained 150 nM RNAP and 2 nM labelled DNA. Samples were analyzed at the indicated times following addition of 3 μM unlabelled DNA. Bands corresponding to free DNA (fDNA) and to specific RNAP–DNA complexes are indicated. On addition of competitor ($t = 0$), the complex fraction was estimated to be 0.8. In the absence of competitor, slower-migrating complexes (C*) that contained several RNAP molecules were seen. All the lanes displayed came from the same gel. (**B**) Time course of RNAP–P_{cr} complex dissociation. Linear fits of data from each of two experimental conditions are shown.

1500-fold excess (3 μM) of unlabelled promoter fragment at $t = 0$, and samples were analyzed at intervals.

We were unable to measure the half-life of the 0°C complexes, since immediately after the addition of the competing unlabelled DNA only ~10% of the labelled promoter DNA remained bound to RNAP, and this percentage barely decreased for the whole 30-min period of competition (not shown). These results reinforce the idea of the heterogeneity of the population of binary 0°C complexes, suggesting that it is mostly composed of very short-lived complexes which are in a rapid equilibrium with free RNAP and promoter. Only a minor fraction of the complexes populated at binding equilibrium at this temperature would be stable, dissociating very slowly into their components. These unstable and stable complexes might correspond, respectively, to the short-footprint and to the extended-footprint complexes that were assumed to coexist at binding equilibrium at 0°C from DNase I footprinting experiments (see above).

On the other hand, most of the binary complexes generated at 37°C resisted a brief exposition to the competitor, as ~80% of the labelled promoter DNA remained bound to RNAP just after the addition of the unlabelled DNA ($t = 0$) (Figure 2A). In addition, dissociation of these RNAP–P_{cr} complexes was very slow (Figure 2B) and appeared to follow a first order kinetic which yielded a dissociation rate constant $k_d = (6.3 \pm 0.7) \times 10^{-5}\,s^{-1}$ ($t_{1/2}$ ~180 min). To discard this low k_d being due to the stoichiometric excess of RNAP (150 nM) over the labelled promoter fragment (2 nM), experiments of dissociation of RNAP–P_{cr} complexes generated at 37°C were also performed by equilibrating 2 nM ^{32}P-labelled DNA and 8 nM RNAP and, at $t = 0$, adding a 100-fold molar excess (200 nM) of unlabelled promoter fragment. In this case (Figure 2B), a $k_d = (6.0 \pm 0.8) \times 10^{-5}\,s^{-1}$ was estimated from the data, which was not significantly different from that obtained when 150 nM RNAP was used. Similar k_d values were obtained when either 50 μg/ml or 150 μg/ml of heparin were added, instead of the unlabelled promoter fragment, to sequester free RNAP from pre-equilibrated binding mixtures containing 2 nM ^{32}P-labelled DNA and 8 nM RNAP (data not shown). These results indicate that the apparent rate of dissociation of RNAP from P_{cr} is not affected by the nature or concentration of the RNAP-quenching agent, which would be thus unable to actively displace RNAP from the complexes, and would only sequester the free enzyme. Estimated k_d values show that the major RNAP–P_{cr} complexes generated at 37°C are very stable, with a half-life of ~3 h. The stability of these complexes is well within the range reported for the RNAP–promoter complexes generated at 37°C (6,33).

CopG prevents binding of RNAP to P_{cr} promoter

To disclose the mechanism of CopG-mediated repression of P_{cr} we first investigated the step at which the repressor prevents transcription from this promoter. To this end, the promoter fragment (2 nM) was equilibrated, at 0°C or 37°C, with excess CopG (120 nM), so that virtual saturation of the operators was obtained. Subsequently, RNAP was added, also at a molar excess (150 nM) relative to the promoter DNA, and incubation continued for 60 more minutes at 0°C or for 30 more minutes at 37°C, prior to DNase I probing of the protein footprints (Figure 3A and B). At either temperature, the footprint pattern of the sample containing both proteins was almost identical to that of the DNA incubated with CopG alone. By decreasing the repressor concentration, the RNAP-specific footprint pattern was increasingly apparent (not shown).

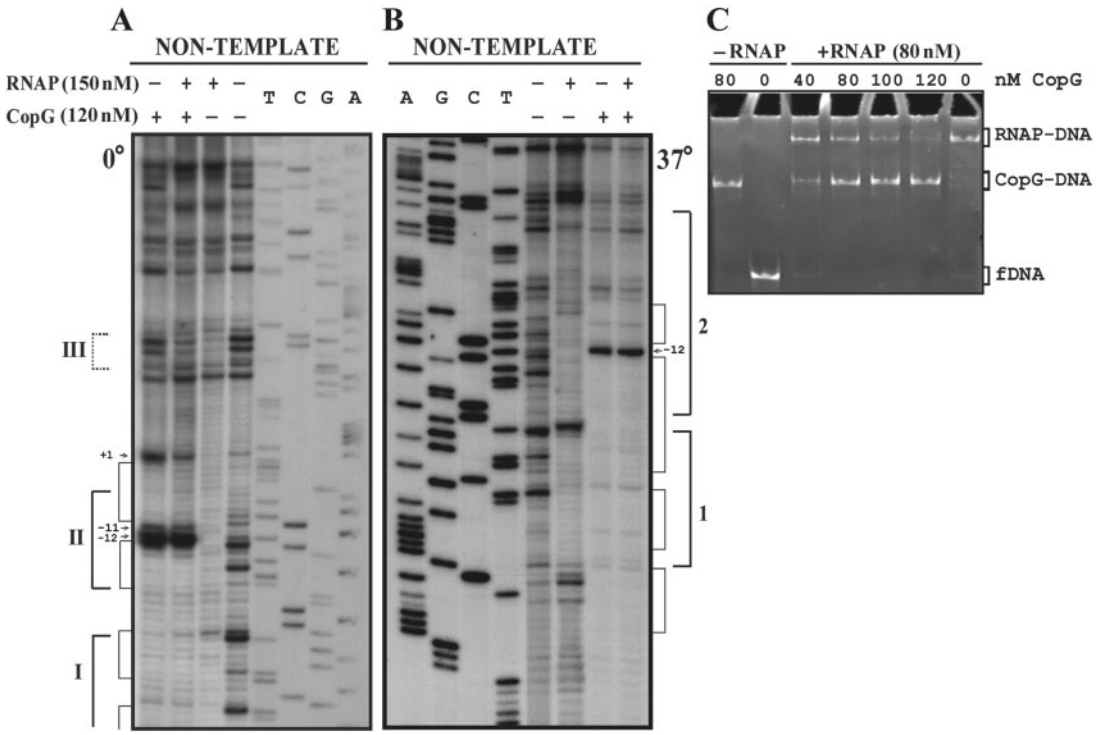


Figure 3. CopG and RNAP compete for binding to the P_{cr} region. (**A**, **B**) DNase I footprinting of binding mixtures containing CopG and/or RNAP. Labelled DNA was incubated at 0°C (A) or 37°C (B) in the presence (+) or in the absence (−) of CopG and RNAP prior to digestion with DNase I. When both proteins were included, CopG was added before RNAP. Cleavage products on the non-template DNA strand are shown, with regions protected by CopG or RNAP indicated by thin or thick brackets, respectively. RNAP footprints are named as in Figure 1B. Enhancements due to CopG binding are indicated by arrows. Dideoxy sequencing reactions on the same DNA are included. (**C**) EMSA analysis of the complexes formed at 37°C in the presence of CopG and/or RNAP. Unlabelled DNA (10 nM) was incubated with the indicated concentrations of CopG prior to addition of RNAP. Samples were treated with heparin before loading onto the gel. Bands corresponding to free DNA (fDNA), and to CopG–DNA and RNAP–DNA complexes are indicated.

These results indicated that CopG was able to prevent subsequent binding of RNAP to P_{cr}. To verify that addition of RNAP to a preformed CopG–DNA complex does not result in the formation of a ternary repressor–DNA–RNAP complex, we analyzed the electrophoretic mobility of DNA samples that were incubated in the absence of CopG or in the presence of increasing concentrations of the repressor prior to the addition of RNAP (Figure 3C). The samples were treated with heparin before loading onto the gel, so that unspecific complexes resulting from the binding of additional molecules of RNAP to non-promoter regions were removed and could not interfere with identification of any existing ternary complex generated by simultaneous binding of both proteins to the operator/promoter region of the DNA. As heparin treatment also removes unstable specific closed complexes, electrophoretic mobility shift assay (EMSA) was only performed with binding mixtures prepared at 37°C. Bands that migrate as the specific binary CopG–DNA or RNAP–DNA complexes, but no super-shifted bands indicative of stable ternary complexes, were seen at all CopG concentrations. Increasing concentrations of the repressor resulted in a decrease of the fraction of DNA bound to RNAP and a concomitant increase of the fraction of CopG–DNA complexes (Figure 3C). Similarly, in experiments in which increasing concentrations of RNAP were added to a CopG–DNA equilibrium mixture, the fraction of RNAP–DNA complex was seen to increase while that of the CopG–DNA complexes decreased (not shown). Thus, the success of CopG in preventing formation of a stable RNAP–P_{cr} complex depends on its own concentration and that of RNAP. These results, together with those of the DNase I footprinting assays performed at 0°C or 37°C in the presence of both proteins, show that binding of CopG and RNAP are mutually exclusive, and that CopG competes with RNAP for binding to the target DNA. RNAP was able to displace DNA-bound CopG, as revealed by the appearance of RNAP–DNA complexes upon addition of the enzyme to binding mixtures in which virtually all the operator sites were saturated with CopG (Figure 3C, compare the samples containing 80 nM CopG, with and without RNAP). A passive displacement of the repressor from its complexes with the operator, resulting from the binding of RNAP to the free DNA present in the CopG equilibrium binding mixture, could account for this observation, provided that CopG–DNA complexes dissociate rapidly on the time scale of the incubation with RNAP.

CopG-operator complexes are unstable

The kinetics of dissociation of CopG-operator complexes at 37°C could only be measured by employing a moderate (50-fold) excess of competing unlabelled DNA (Figure 4). When a higher excess (500- or 1500-fold) was used, no

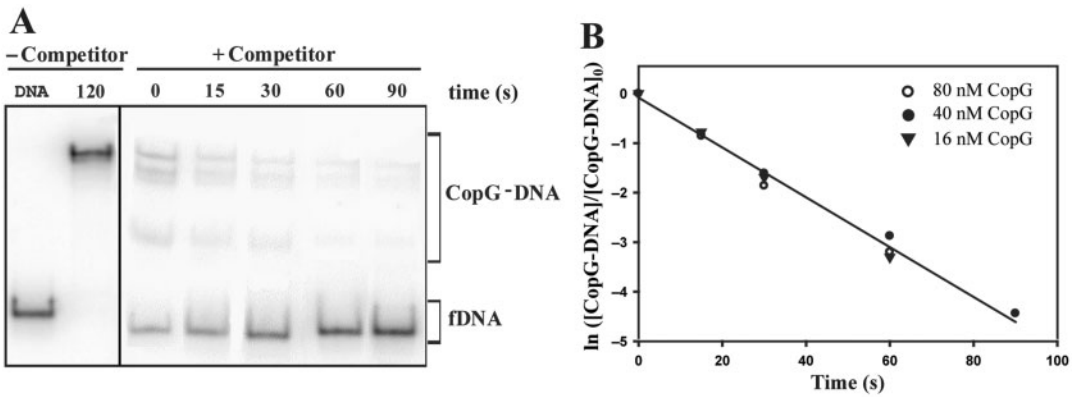


Figure 4. Kinetic of dissociation of CopG from its operator at 37°C. (**A**) EMSA analysis of the stability of the CopG–DNA complexes. Dissociation of complexes between CopG (40 nM) and the labelled DNA was initiated by addition of a 50-fold excess of unlabelled DNA ($t = 0$), and samples were analyzed at the indicated times. The sum of the various CopG–DNA complexes (in brackets) was used to analyze the time course of the fraction of labelled DNA complexed to CopG. Samples of free DNA (fDNA) and of the equilibrium mixture without competitor were also loaded at $t = 120$. All the lanes displayed came from the same gel. (**B**) Time course of CopG–DNA complex dissociation. Data from three independent experiments, each performed at the indicated CopG concentration, are included. The k_d values estimated were $(5.5 \pm 0.1) \times 10^{-2}\,s^{-1}$ (16 nM CopG), $(4.8 \pm 0.2) \times 10^{-2}\,s^{-1}$ (40 nM CopG), and $(5.3 \pm 0.7) \times 10^{-2}\,s^{-1}$ (80 nM CopG). The solid line is the linear fit of all data.

DNA complexed to CopG could be detected immediately on addition of the unlabelled DNA (not shown). Thus, the lifetime of CopG-operator complexes seems to depend strongly on the concentration of competing DNA, which suggests that 'direct transfer' of CopG to another DNA molecule contributes to the dissociation mechanism of these complexes. DNA concentration-dependent dissociation has been shown for the *lac* repressor and CAP proteins complexed to their target DNAs (34,35). Data for dissociation kinetic arose from three independent experiments, each performed at a different CopG concentration and thus displaying a distinct distribution of free and complexed DNA in the equilibrium mixture prior to the addition of the competing DNA. All of them gave similarly high dissociation rate constants (Figure 4). When data of all three experiments were analyzed conjointly according to Equation (1) for a pseudo first-order process, the apparent dissociation rate constant was estimated to be $(5.0 \pm 0.3) \times 10^{-2}\,s^{-1}$ (Figure 4), which corresponded to a $t_{1/2} \sim 14$ s. The low stability of the CopG-operator complexes at 37°C could account for the observed displacement of DNA-bound CopG by RNAP (Figure 3C), as the addition of the polymerase to a previously equilibrated CopG–DNA binding mixture would passively displace the equilibrium in the direction of dissociation of the CopG–DNA complexes, by sequestering free DNA.

CopG dislodges RNAP stably bound to P_{cr} promoter

Evidences of the ability of CopG to displace RNAP stably bound to the P_{cr} promoter were obtained by inverting the order in which the repressor and the polymerase were added to the target DNA. In the absence of CopG, RNAP formed very stable complexes with P_{cr} at 37°C (Figure 2). When preformed RNAP–P_{cr} complexes were challenged for 10 min with increasing concentrations of CopG, a decline of the RNAP-specific complexes was observed which was paralleled by an increase in the amount of CopG–DNA complexes (Figure 5A and B). The CopG–DNA complexes generated in the presence of RNAP exhibited the same electrophoretic mobility as those produced in the absence of the polymerase at identical repressor concentrations (Figure 5A and B). CopG-dependent decrease of the RNAP–P_{cr} complexes was observed irrespective of whether the concentration of RNAP used in the assays was 150 nM (Figure 5A) or 8 nM (Figure 5B). Almost complete removal of the RNAP–P_{cr} complexes was obtained after treatment at the highest CopG concentrations (Figure 5A and B), even though the 10-min incubation time is quite short relative to the half-life of these complexes (~180 min). Therefore, it appeared of interest to analyze the time course of the CopG-mediated displacement of RNAP from P_{cr} by challenging the specific RNAP–P_{cr} complexes with CopG for different times. Several independent experiments were performed over a range of concentrations of CopG, RNAP and heparin (four of these experiments are shown in panels A, B, C and D of Figure 6). Each experiment included control samples in which only heparin (instead of heparin followed by CopG) was added at $t = 0$. In every experiment, the CopG-containing samples exhibited repressor-concentration-dependent decreases in the fraction of RNAP–P_{cr} complexes, compared with controls (Figure 6 and Supplementary Tables 1, 2, 3 and 4). Maximal extent of the CopG-mediated RNAP displacement was reached within 1 min of incubation with the repressor, while longer incubations did not result in any further significant decrease in the fraction of RNAP–P_{cr} complexes relative to the corresponding control samples (Figure 6 and Supplementary Tables 1, 2, 3 and 4). It is worth noting that the experiment in Figure 6D was carried out employing the *E. coli* RNAP holoenzyme from USB and two different heparin concentrations. As can be seen, the use of a heparin concentration more than three times greater than that required to compete unstable RNAP–DNA complexes did not significantly affect either the stability of the RNAP–P_{cr} specific complexes in the absence of CopG or the level and time course

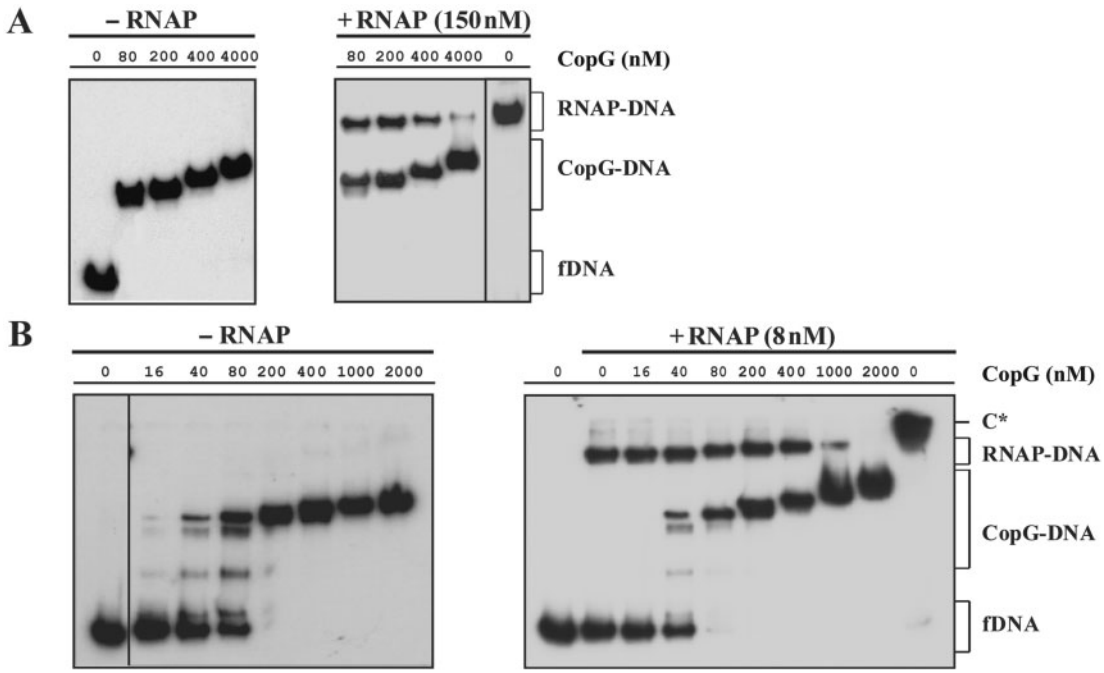


Figure 5. CopG-concentration-dependent dissociation of the RNAP–P_{cr} complexes at 37°C. RNAP (at the concentrations indicated in **A** and in **B**) and DNA (2 nM) were equilibrated at 37°C and then treated with heparin (150 μg/ml in A, and 10 μg/ml in B) before adding different amounts of CopG. Incubation of the mixtures continued for 10 more minutes before loading onto the gel. CopG–DNA complexes formed in the absence of RNAP were also analyzed. Bands corresponding to free DNA (fDNA) and to RNAP–DNA and CopG–DNA complexes are indicated. At high CopG concentrations (>200 nM), complexes migrating slower than the specific complex generated by binding of four CopG dimers to the operator DNA (17) were observed. These slowly-migrating complexes contain additional repressor molecules nucleated from the operator (39). All the lanes displayed came either from the same gel or from gels prepared and run in parallel. Only one of two independent experiments yielding identical results is shown in each panel. Note that in the absence of heparin, slowly migrating RNAP–DNA complexes that contained several RNAP molecules were seen (C* in panel B).

of the CopG-mediated displacement of RNAP from the promoter (Figure 6D and Supplementary Table 4). The time-course experiments also showed that CopG-mediated dissociation of the RNAP–P_{cr} complexes occurs regardless the concentration and the source of RNAP employed, which allows us to rule out any artefacts due to a particular commercial holoenzyme preparation.

Taken together, the above results show that CopG actively replaces the polymerase bound to the promoter/operator region. The CopG-mediated displacement of RNAP from P_{cr} depended on the repressor concentration and constituted a rapid process that took place in less than 1 min after addition of the repressor (Figures 5 and 6, and Supplementary Tables 1, 2, 3 and 4).

CopG and RNAP contact the promoter/operator region by different, though overlapping, sides of the DNA double helix

To analyze the mechanism by which CopG displaces RNAP from P_{cr} we determined the precise contacts of both proteins with the DNA backbone of the promoter/operator region (Figure 7 and Supplementary Figures S3 and S4).

HO• footprinting of the CopG–DNA specific complexes generated at 37°C revealed five protected regions on each strand (a, b, c, d and e; Supplementary Figure S3), which basically matched the footprints observed upon binding of the protein at 20°C (15). CopG footprints showed that the protein binds to one face of the DNA helix (15), and delimit four regions of the major groove (Figure 8) whose bases are specifically contacted by CopG dimers [(17); unpublished data]. The two central CopG-binding sites correspond to the left (LSE) and right (RSE) halves of the SE, whose atomic structure in complex with CopG has been solved (16). The two additional binding sites, hereafter named Left Arm (LA) and Right Arm (RA), are located one helical turn apart from the LSE and the RSE, respectively (Figure 8).

HO• probing of the heparin-resistant RNAP–P_{cr} complexes generated at 37°C gave a footprint pattern that corresponded to extended binary complexes (Figure 7 and Supplementary Figure S4) and was similar to those observed for open complexes in other RNAP/promoter systems (25,31,36). Upstream of the −10 region of P_{cr} the periodicity of the RNAP footprints indicated that the polymerase contacted the DNA backbone by one face of the double helix, whereas downstream from this region the contacts involved the entire DNA helix (Figure 8). The large downstream footprint interrupted only around position +1 on the template strand at a few sugar-phosphate bonds which were sensitive or even hypersensitive towards HO• (Supplementary Figures S2 and S4). This strand-specific unprotected region around

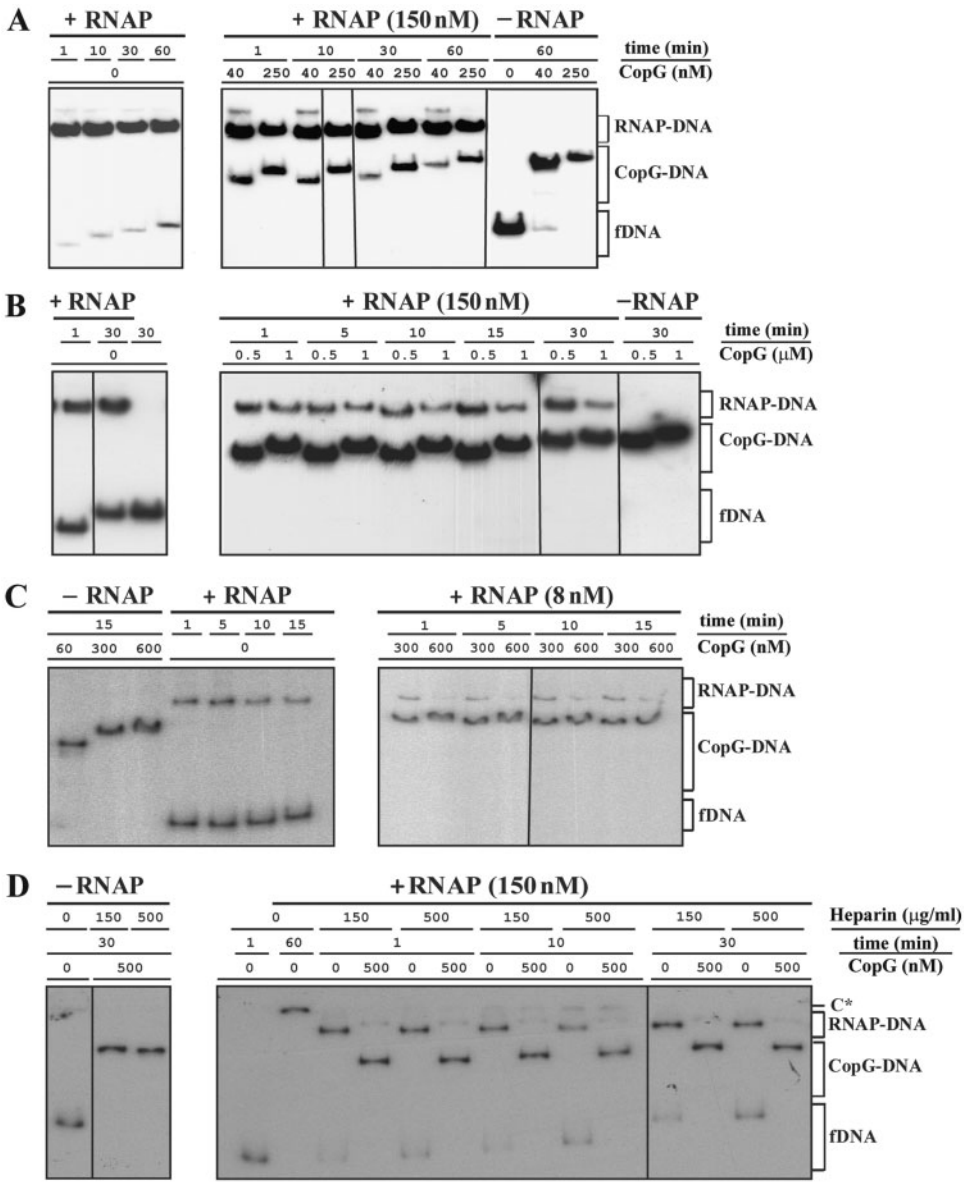


Figure 6. CopG-mediated rapid dissociation of the RNAP–P_{cr} complexes at 37°C. (**A**, **B**, **C**, **D**) EMSA from four independent time-course experiments performed over a range of CopG and RNAP concentrations. RNAP holoenzyme was purchased from Boehringer (panels A, B and C) or from USB (panel D). DNA (2 nM) and RNAP at the indicated concentrations were equilibrated at 37°C and then treated at $t = 0$ with heparin (150 μg/ml in A and B, 10 μg/ml in C, and either 150 μg/ml or 500 μg/ml in D) followed by different concentrations of CopG. Incubation of the mixtures continued for several times before loading onto the gel. CopG–DNA complexes formed in the absence of RNAP were also analyzed. Bands corresponding to free DNA (fDNA) and to RNAP–DNA and CopG–DNA complexes are indicated. At high CopG concentrations (>200 nM), CopG–DNA complexes migrating slower than the specific complex generated by binding of four repressor dimers were observed. All the lanes shown in each experiment came either from the same gel or from gels prepared and run in parallel. Note that in the absence of heparin, slowly migrating RNAP–DNA complexes that contained several RNAP molecules were observed (C* in panel D). The percentages of RNAP–DNA and CopG–DNA complexes quantified from the EMSA displayed in panels A, B, C and D, are shown in Supplementary Tables 1, 2, 3 and 4, respectively.

the transcription start site could be considered as a hallmark of the open complexes (25,27,36), which most likely populate the RNAP–P_{cr} complexes generated at 37°C. Upstream from the −10 region of P_{cr}, contacts of CopG with the DNA backbone were shifted on the double helix by some 80° (about 0.23 helix turns on average) relative to those of RNAP, although some sugar-phosphate bonds were protected by both proteins (Figure 8). Hence, CopG and RNAP interact with the promoter/operator region on different faces of the DNA.

The RSE of the operator DNA is the primary binding site of CopG

To know the relevance of the different elements constituting the operator of CopG, we analyzed the effect of changing the nucleotide sequence of each of the four CopG-binding sites on the affinity of the protein. To this end, we measured the binding affinity of CopG for mutant operators LSE^-, RSE^-, LA^- and RA^- relative to the wild-type operator (see the scheme of the mutants in

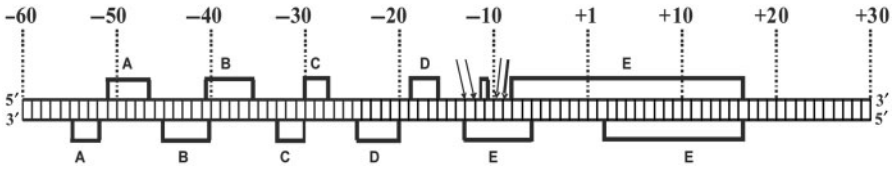


Figure 7. Summary of HO• footprints of the RNAP–P_{cr} complexes at 37°C. The schematized protection pattern corresponds to the footprinting experiment in Supplementary Figure S4. Promoter positions relative to the transcription start site are indicated. Protections are shown by brackets. Protections denoted by the same letter in both strands lie across the minor groove of the DNA. Enhancements are shown by arrows.

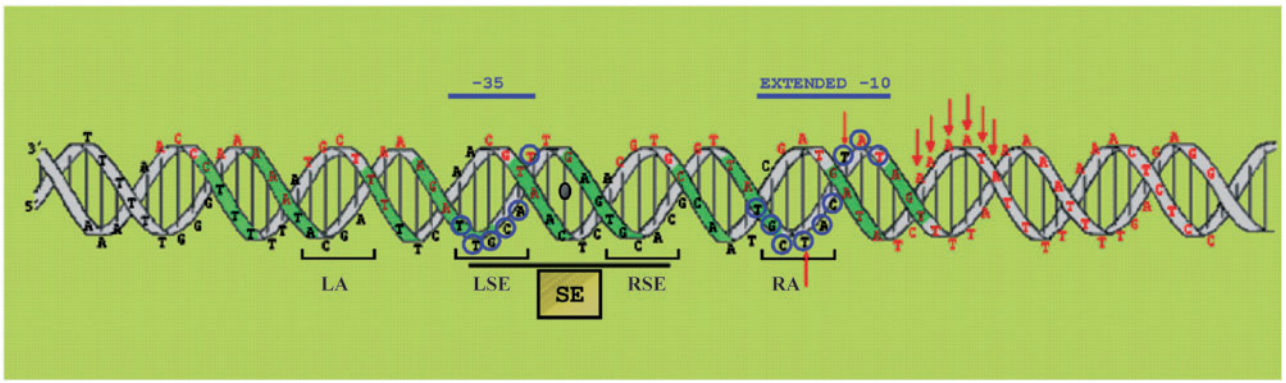


Figure 8. Schematic representation of the contacts of CopG and RNAP with the DNA backbone of the P_{cr} promoter region at 37°C. RNAP contacts are denoted by red letters. Red arrows indicate positions in the RNAP–DNA complex with enhanced backbone cleavage. Contacts by CopG are shadowed in green. The SE is boxed. The binding sites of CopG dimers are indicated with brackets. The right half of the SE (RSE) is the CopG primary site. The −35 and extended −10 sequences of P_{cr} are denoted by lines and encircled letters in blue colour.

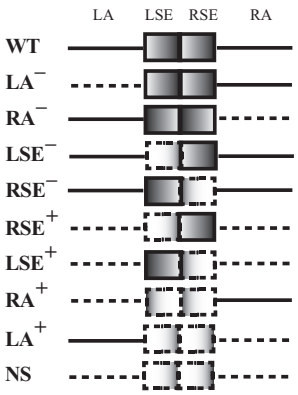
Table 1 and the experimental design in Supplementary Materials and Methods section and Figure S5). Substitution of the RSE by an unspecific sequence affected binding of CopG most severely, resulting in a ~300-fold reduction in the affinity of the protein for the RSE^- mutant operator (Table 1). In contrast, a much lower reduction (by only ~6-fold) in binding affinity was obtained when the LSE was replaced by the unspecific sequence (Table 1). Changing the sequence of either the LA or the RA also caused differential effects on the affinity of CopG binding: whereas replacement of the LA reduced the affinity by ~2.5-fold, substitution of the RA had the same effect (a 6–7-fold reduction in the affinity) as changing the LSE (Table 1). These results indicated that the RSE contributed most to the high affinity of binding of CopG to its operator, while the LA had the lowest contribution. To verify whether the RSE was the preferred CopG-binding site, the relative affinity of the protein for mutant operators containing a single binding site was also measured (Table 1 and Supplementary Figure S5). CopG exhibited detectable specific binding affinity exclusively for the RSE^+ operator, which only harboured the wild-type RSE. The relative affinity of CopG for the other single-binding-site operator mutants could not be measured, as was also the case for the non-specific (NS) oligonucleotide lacking the four CopG-binding sites (Table 1 and Supplementary Figure S5). These results show that the affinity of binding of CopG to the LSE^+, LA^+ and RA^+ operator mutants is very low and most likely approaches non-specific levels. Remarkably, the CopG affinity for the RSE^+ operator was ~2.5-fold greater than the affinity of the protein for the RSE^- mutant operator (Table 1), therefore showing that the presence of the sole RSE is more important for the binding affinity of the protein than the presence of the other three sites together. The overall results obtained from the analysis of the affinity of CopG for operator variants allow us to conclude that the RSE is the binding site preferred by CopG within the operator, that is to say, the RSE is the primary binding site of the protein. Moreover, the RSE is the only binding site whose presence guarantees specific binding of CopG. Since the RSE is required for specific and high-affinity complex formation by CopG, it appears that accessibility of this binding site is an essential factor for the binding of this protein to its operator.

DISCUSSION

Plasmid pMV158 possesses a very promiscuous replicon that functions in *E. coli* (37). This implies that P_{cr} promoter, which directs transcription of the *repB* gene essential for plasmid replication, is recognized by RNAP from *E. coli*. In fact, transcription from P_{cr} by *E. coli* RNAP as well as CopG-mediated repression of it have been reported both *in vivo* and *in vitro* (15), although the bases underlying these processes had not been investigated. Here, we have characterized the complexes formed by RNAP at P_{cr} and analyzed the mechanism by which CopG represses transcription from this promoter.

As deduced from the contradictory story with Lac repressor (6), elucidation of the precise step at which a given repressor protein inhibits the reversible process of formation of a transcriptionally active open complex requires the previous knowledge of the dissociation rate

Table 1. CopG binding affinity to different operator variants

Target DNA[a]	Relative affinity[b]
	
WT	1
LA^-	0.382 ± 0.046
RA^-	0.150 ± 0.014
LSE^-	0.162 ± 0.022
RSE^-	0.0034 ± 0.0005
RSE^+	0.0082 ± 0.0014
LSE^+	< 0.0001
RA^+	< 0.0001
LA^+	< 0.0001
NS	< 0.0001

[a]The four binding sites of the operator (LA, LSE, RSE and RA) are indicated on the schematic representation of the wild-type target DNA (WT). Continuous lines denote wild-type sequence regions. Dashed lines represent DNA regions with substituted sequence. The names of the mutant operators are indicated on the left.

[b]The affinities of binding of CopG to the 55-bp operator variants relative to the 239 wild-type DNA fragment are shown. The number of samples independently analyzed to calculate each relative binding affinity varied between 17 and 33. The calculated relative affinities were normalized for the value of the ratio between the affinity of CopG for the 55-bp wild-type oligonucleotide and the affinity of the protein for the 239-bp wild-type operator DNA fragment (0.85 ± 0.04).

constants of the operator/repressor and promoter/RNAP complexes. At promoter P_{cr}, typical unstable short closed and stable extended open complexes were generated upon binding of RNAP at 0°C and 37°C, respectively (Figures 1 and 2). The stability of the RNAP–P_{cr} complexes at 37°C ($t_{1/2} \sim 3$ h; Figure 2) contrasted with the short half-life ($t_{1/2} \sim 14$ s) of the complexes generated by binding of CopG to the promoter/operator region at this temperature (Figure 4). Hence, on the time scale of our experiments at 37°C, CopG–DNA complexes dissociated rapidly whereas RNAP–P_{cr} complexes hardly dissociated.

Clues to the mechanism of CopG-mediated repression of transcription from P_{cr} were provided from competition assays in which the order of addition of polymerase and repressor was inverted: (i) no ternary specific complexes were observed in any case (Figure 3C), which evidences mutually exclusive binding of CopG and RNAP to the promoter/operator region; (ii) CopG induced disassembly of preformed stable RNAP–P_{cr} complexes (Figures 5 and 6, and Supplementary Tables 1, 2, 3 and 4); and (iii) in the presence of CopG and RNAP, the apparent equilibrium extent of formation of RNAP–P_{cr} complexes depended on the concentration of both proteins (Figures 3 and 5). Therefore, CopG seems to repress transcription initiation from P_{cr} by acting at two different steps in the process of formation of an open complex. First, CopG competes with RNAP for binding to the same region of DNA, where promoter and operator overlap. At this stage, CopG bound to its target DNA would hinder the access of RNAP to the promoter region by a steric exclusion mechanism, which constitutes a classic model of repressor action. The strong DNA bend induced by CopG toward the repressor–DNA interface (16) might play a role in this mechanism as well. Secondly, CopG is able to displace RNAP stably bound to P_{cr}, which constitutes an unexpected ability with no obvious explanation. We propose a model for the CopG-induced disassembly of the stable RNAP–P_{cr} complexes which takes into account the small size of CopG (a dimeric protein of 45 amino acids per subunit), the sequential and cooperative binding of CopG to the four sites that constitute the operator, and the accessibility of the CopG primary binding site in the RNAP–P_{cr} complexes. Specific CopG–operator complexes result from cooperative binding of four CopG dimers on the same face of the DNA helix (Figure 8). The repressor interacts specifically through the major groove of the DNA with four sites separated one helical turn from each other (17). The two central sites correspond to the LSE and RSE elements of the operator (Figure 8). Each of the two outer sites (LA or RA) is located one helical turn apart from either central site. By employment of operator mutants that either lack or harbour only one of these binding sites, we have shown that the RSE is the primary binding site of CopG (Table 1 and Supplementary Figure S5) and hence it is, most likely, the first site occupied during the sequential and cooperative binding of the repressor. This primary CopG-binding site could be accessible to the repressor even in the stable RNAP–P_{cr} complexes generated at 37°C since at the RSE region there are only two phosphate groups contacted by both proteins (Figure 8), and there are no expected contacts of RNAP with specific bases of the DNA. This situation is unique for the primary binding site of CopG, because the LSE and RA sites where subsequent repressor molecules would bind contain, respectively, the −35 and the −10 hexamers of P_{cr}, as well as a higher number of phosphate groups contacted by CopG and RNAP (Figure 8). Thus, the proposed sequence of events in the model for CopG-induced disassembly of the stable RNAP–P_{cr} complexes would be as follows: a repressor dimer reaches its primary binding site at the right half of the SE and establishes contacts with specific bases of the DNA through the major groove. This preliminary CopG–operator interaction might change the conformation of the DNA, thus weakening the contacts between RNAP and the nucleic acid. If repressor concentration is high enough, binding of CopG to the primary site may promote rapid successive binding to the other sites in the operator, so that CopG would displace RNAP from the promoter (Figure 8). By acting at two different steps during open complex formation, repression by CopG does not consist in a simple mechanism of competition between free molecules of RNAP and repressor for binding to the same target DNA, since CopG can also establish interactions with its operator when RNAP is already bound to the P_{cr} promoter. Repression of transcription initiation at two different steps might allow a fine-tuning of the expression from P_{cr} in response to the intracellular concentration of CopG. Hence, the proposed dual repression mechanism could be appropriate for a control system involved in regulation of plasmid

replication, as variations in the intracellular conditions would be rapidly sensed due to the short half-life of the CopG/operator complexes and to the ability of CopG to dislodge the bound RNAP.

Repression of transcription initiation by disassembly of the open complex seems to be an uncommon mechanism, which has only been reported previously for the *E. coli* repressor IclR (38). IclR represses expression of the *aceBAK* operon (involved in the glyoxylate bypass) through two different mechanisms, each acting at a distinct step of the open complex formation process. First, IclR binds to its primary site (IclR box II), located between −52 and −19 of the *aceB* promoter, thus interfering with binding of RNAP to this promoter. Second, at higher repressor concentrations, IclR binds to its secondary site (IclR box I), located between −125 and −99 of the *aceB* promoter, even after formation of an open complex on this promoter, and induces disassembly of RNAP from the promoter. This latter mechanism of transcription repression is based on the establishment of direct interactions between the C-terminal domain of RNAP α subunit and IclR bound to its secondary binding site (38).

CopG- and IclR-mediated repression of transcription from P_{cr} and *aceB* promoters, respectively, displays both similarities and differences. CopG and IclR can act at two distinct steps of the open complex formation, namely by competing with RNAP for binding to the DNA or by disassembling the open complex at the respective promoter. However, whereas repression by IclR involves two different mechanisms (steric exclusion of, and interaction with, RNAP) and two different binding sites (boxes II and I), cooperative binding of four CopG molecules to its single operator would be responsible for the dual repression mechanism reported here. From this, a picture emerges of CopG being the minimal protein element able to bind specifically and cooperatively to the operator in such a way that it not only prevents the binding of the RNAP, but also displaces efficiently the polymerase bound to the regulated promoter.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

Thanks are due to M.T. Alda for purification of CopG; to members of our lab for helpful discussions and to A.M.L. Jones for correcting the English throughout the manuscript.

FUNDING

Ministerio de Ciencia e Innovación [grants BFU2004-00687/BMC and BFU2007-63575 to G.d.S.]; [grant INTERMODS-CSD2008-00013 to M.E.]; A.M.H.A. by [grant BFU2005-03911 to Prof. R. Díaz-Orejas]. Funding for open access charge: Grant BFU2007-63575 to G.d.S. from Ministerio de Ciencia e Innovación.

Conflict of interest statement. None declared.

REFERENCES

1. Rojo,F. (1999) Repression of transcription initiation in bacteria. *J. Bacteriol.*, **181**, 2987–2991.
2. Rojo,F. (2001) Mechanisms of transcriptional repression. *Curr. Opin. Microbiol.*, **4**, 145–151.
3. Hawley,D.K., Johnson,A.D. and McClure,W.R. (1985) Functional and physical characterization of transcription initiation complexes in the bacteriophage λ O_R region. *J. Biol. Chem.*, **260**, 8618–8626.
4. Bertrand-Burggraf,E., Hurstel,S., Daune,M. and Schnarr,M. (1987) Promoter properties and negative regulation of the *uvrA* gene by the LexA repressor and its amino-terminal DNA binding domain. *J. Mol. Biol.*, **193**, 293–302.
5. Majors,J. (1975) Initiation of *in vitro* mRNA synthesis from the wild-type *lac* promoter. *Proc. Natl Acad. Sci. USA*, **72**, 4394–4398.
6. Schlax,P.J., Capp,M.W. and Record,M.T. (1995) Inhibition of transcription initiation by *lac* repressor. *J. Mol. Biol.*, **245**, 331–350.
7. Nieto,C., Puyet,A. and Espinosa,M. (2001) MalR-mediated Regulation of the *Streptococcus pneumoniae malMP* Operon at Promoter PM. Influence of a proximal divergent promoter region and competition between MalR and RNA polymerase proteins. *J. Biol. Chem.*, **276**, 14946–14954.
8. Choy,H.E., Park,S.-W., Aki,T., Parrack,P., Fujita,N., Ishihama,A. and Adhya,S. (1995) Repression and activation of transcription by Gal and Lac repressors: involvement of alpha subunit of RNA polymerase. *EMBO J.*, **14**, 4523–4529.
9. Kuhnke,G., Theres,C., Fritz,H.J. and Ehring,R. (1989) RNA polymerase and *gal* repressor bind simultaneously and with DNA bending to the control region of the *Escherichia coli* galactose operon. *EMBO J.*, **8**, 1247–1255.
10. Monsalve,M., Mencía,M., Rojo,F. and Salas,M. (1996) Activation and repression of transcription at two different phage ϕ29 promoters are mediated by interaction of the same residues of regulatory protein p4 with RNA polymerase. *EMBO J.*, **15**, 383–391.
11. Schneider,R., Travers,A., Kutateladze,T. and Muskhelishvili,G. (1999) A DNA architectural protein couples cellular physiology and DNA topology in *Escherichia coli*. *Mol. Microbiol.*, **34**, 953–964.
12. Schröder,O. and Wagner,R. (2000) The bacterial DNA-binding protein H-NS represses ribosomal RNA transcription by trapping RNA polymerase in the initiation complex. *J. Mol. Biol.*, **298**, 737–748.
13. Shin,M., Song,M., Rhee,J.H., Hong,Y., Kim,Y.-J., Seok,Y.-J., Ha,K.-S., Jung,S.-H. and Choy,H.E. (2005) DNA looping-mediated repression by histone-like protein H-NS: specific requirement of $E\sigma^{70}$ as a cofactor for looping. *Genes Dev.*, **19**, 2388–2398.
14. Xu,J. and Koudelka,G.B. (2001) Repression of transcription initiation at 434 P_R by 434 repressor: effects on transition of a closed to an open promoter complex. *J. Mol. Biol.*, **309**, 573–587.
15. del Solar,G., Pérez-Martín,J. and Espinosa,M. (1990) Plasmid pLS1-encoded RepA protein regulates transcription from *repAB* promoter by binding to a DNA sequence containing a 13-base pair symmetric element. *J. Biol. Chem.*, **265**, 12569–12575.
16. Gomis-Rüth,F.X., Solá,M., Acebo,P., Párraga,A., Guasch,A., Eritja,R., González,A., Espinosa,M., del Solar,G. and Coll,M. (1998) The structure of plasmid-encoded transcriptional repressor CopG unliganded and bound to its operator. *EMBO J.*, **17**, 7404–7415.
17. del Solar,G., Hernández-Arriaga,A.M., Gomis-Rüth,F.X., Coll,M. and Espinosa,M. (2002) A genetically economical family of plasmid-encoded transcriptional repressors involved in control of plasmid copy number. *J. Bacteriol.*, **184**, 4943–4951.
18. Phillips,S.E.V. (1991) Specific β-sheet interactions. *Curr. Opin. Struct. Biol.*, **1**, 89–98.
19. Raumann,B.E., Rould,M.A., Pabo,C.O. and Sauer,R.T. (1994) DNA recognition by beta-sheets in the Arc repressor-operator crystal structure. *Nature*, **367**, 754–757.
20. Berggrun,A. and Sauer,R.T. (2001) Contributions of distinct quaternary contacts to cooperative operator binding by Mnt repressor. *Proc. Natl Acad. Sci. USA*, **98**, 2301–2305.

21. Somers,W.S. and Phillips,S.E.V. (1992) Crystal structure of the *met* repressor-operator complex at 2.8 Å resolution reveals DNA recognition by β-strands. *Nature*, **359**, 387–393.
22. Smith,T.L. and Sauer,R.T. (1995) P22 Arc repressor: role of cooperativity in repression and binding to operators with altered half-site spacing. *J. Mol. Biol.*, **249**, 729–742.
23. Smith,T.L. and Sauer,R.T. (1996) Role of operator subsites in Arc repression. *J. Mol. Biol.*, **264**, 233–242.
24. Gomis-Rüth,F.X., Solá,M., Pérez-Luque,R., Acebo,P., Alda,M.T., González,A., Espinosa,M., del Solar,G. and Coll,M. (1998) Overexpression, purification, crystallization and preliminary X-ray diffraction analysis of the pMV158-encoded plasmid transcriptional repressor protein CopG. *FEBS Lett.*, **425**, 161–165.
25. Schickor,P., Metzger,W., Werel,W., Lederer,H. and Heumann,H. (1990) Topography of intermediates in transcription initiation of *E. coli*. *EMBO J.*, **9**, 2215–2220.
26. Maxam,A.M. and Gilbert,W. (1980) Sequencing end-labelled DNA with base-specific chemical cleavages. *Methods Enzymol.*, **65**, 499–559.
27. Record,M.T., Reznikoff,W.S., Craig,M.L., McQuade,K.L. and Schlaz,P.J. (1996) In Neidhart,F.C., Curtiss III,R., Ingraham,J.L., Lin,E.C.C., Low,K.R., Magasanik,B., Reznikoff,W.S., Riley,M., Schaetchter,M. and Umbarger,H.E. (eds), *Escherichia coli and Salmonella: Cellular and Molecular Biology*. 2nd edn. ASM Press, Washington, DC, pp. 792–820.
28. Kirkegaard,K., Buc,H., Spassky,A. and Wang,J.C. (1983) Mapping of single-stranded regions in duplex DNA at the sequence level: single-strand-specific cytosine methylation in RNA polymerase-promoter complexes. *Proc. Natl Acad. Sci. USA*, **80**, 2544–2548.
29. Straney,D.C. and Crothers,D.M. (1985) Intermediates in transcription initiation from the *E. coli lac* UV5. *Cell*, **43**, 449–459.
30. Craig,M.L., Tsodikov,O.V., McQuade,K.L., Schlax,P.E., Capp,M.W., Saecker,R.M. and Record,M.T. (1998) DNA footprints of the two kinetically significant intermediates in formation of an RNA polymerase-promoter open complex: evidence that interactions with start site and downstream DNA induce sequential conformational changes in polymerase and DNA. *J. Mol. Biol.*, **283**, 741–756.
31. Li,X.-Y. and McClure,R.W. (1998) Characterization of the closed complex intermediate formed during transcription initiation by *Escherichia coli* RNA polymerase. *J. Biol. Chem.*, **273**, 23549–23557.
32. Cowing,D.W., Mecsas,J., Record,M.T. Jr. and Gross,C.A. (1989) Intermediates in the formation of the open complex by RNA polymerase holoenzyme containing the sigma factor σ^{32} at the groE promoter. *J. Mol. Biol.*, **210**, 521–530.
33. Roe,J.-H., Burgess,R.R. and Record,M.T. Jr. (1984) Kinetics and mechanism of the interaction of *Escherichia coli* RNA polymerase with the λP_R promoter. *J. Mol. Biol.*, **176**, 495–521.
34. Fried,M.G. and Liu,G. (1994) Molecular sequestration stabilizes CAP-DNA complexes during polyacrylamide gel electrophoresis. *Nucleic Acids Res.*, **22**, 5054–5059.
35. Vossen,K.M. and Fried,M.G. (1997) Sequestration stabilizes *lac* repressor-DNA complexes during gel electrophoresis. *Anal. Biochem.*, **245**, 85–92.
36. Craig,M.L., Suh,W.-C. and Record,M.T. Jr. (1995) HO• and DNase I probing of Eσ^{70} RNA polymerase-λP_R promoter open complexes: Mg^{2+} binding and its structural consequences at the transcription start site. *Biochemistry*, **34**, 15624–15632.
37. Lacks,S.A., López,P., Greenberg,B. and Espinosa,M. (1986) Identification and analysis of genes for tetracycline resistance and replication functions in the broad-host-range plasmid pLS1. *J. Mol. Biol.*, **192**, 753–765.
38. Yamamoto,K. and Ishihama,A. (2003) Two different modes of transcription repression of the *Escherichia coli* acetate operon by IclR. *Mol. Microbiol.*, **47**, 183–194.
39. Costa,M., Solá,M., del Solar,G., Eritja,R., Hernández-Arriaga,A.M., Espinosa,M., Gomis-Rüth,F.X. and Coll,M. (2001) Plasmid transcriptional repressor CopG oligomerises to render helical superstructures unbound and in complexes with oligonucleotides. *J. Mol. Biol.*, **310**, 403–417.

終わり